

STIC-Biotech/ChemLib

170010

From: Walicka, Malgorzata  
Sent: Saturday, October 29, 2005 10:56 AM  
To: STIC-Biotech/ChemLib

Please search SEQ ID NO:1 and 16 in 09/774,954. Please use also pending data.

Thank you.

Malgorzata A. Walicka, Ph.D.  
Patent Examiner  
Art Unit 1652, Recombinant Enzymes  
USPTO, Remsen Building, Room 2C76  
400 Dulany St.  
Alexandria, VA 22313  
Mail Room 2C70  
Tel. (571) 272-0944, fax (571) 273-0944

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: 11/4/05  
Date completed: 11/9/05  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# 2 AA#: \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: 04  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:53:44 ; Search time 6830.38 Seconds  
(without alignments)  
10740.434 Million cell updates/sec

Title: US-09-774-954-1  
Perfect score: 1514  
Sequence: 1 ATGCCCGCGGCTCTCGGA.....TTTTGAAATCAATGAATTC 1514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pt.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description          |
|------------|--------|-------------|--------|-------|----------------------|
| 1          | 1514   | 100.0       | 1514   | 6     | AR163451             |
| 2          | 1514   | 100.0       | 1514   | 6     | BD103188             |
| 3          | 1512.4 | 99.9        | 5249   | 9     | AF375884 Homo sapi   |
| 4          | 1512.4 | 99.9        | 11284  | 6     | BD103190 O-fucosyl   |
| 5          | 1510.8 | 99.8        | 5189   | 9     | D80002 Homo sapien   |
| 6          | 1509.8 | 99.7        | 5266   | 6     | AX780153             |
| 7          | 1507.6 | 99.6        | 4560   | 9     | HSM803330            |
| 8          | 1447.4 | 95.6        | 5009   | 6     | AR163453             |
| 9          | 1447.4 | 95.6        | 5009   | 6     | BD103191             |
| 10         | 1163.4 | 76.8        | 1300   | 6     | AR163452             |
| 11         | 1163.4 | 76.8        | 1300   | 6     | BD103189             |
| 12         | 1100   | 72.7        | 1100   | 6     | AR163457             |
| 13         | 1100   | 72.7        | 1100   | 6     | BD103196             |
| 14         | 1096.4 | 72.4        | 1167   | 6     | CQ727777             |
| 15         | 1086.8 | 71.8        | 1167   | 9     | AJ781500             |
| 16         | 916.2  | 60.5        | 1176   | 4     | AY344580 Bos tauru   |
| 17         | 908.2  | 60.0        | 1176   | 4     | AJ781503             |
| 18         | 901.6  | 59.6        | 1134   | 4     | SSC567917 Sus scrofa |
| 19         | 889.6  | 58.8        | 1188   | 10    | AJ781499 Rattus no   |

|    |       |      |        |    |                    |
|----|-------|------|--------|----|--------------------|
| 20 | 889.6 | 58.8 | 1302   | 10 | AF375885           |
| 21 | 889.6 | 58.8 | 1529   | 10 | BC046295           |
| 22 | 752.8 | 49.7 | 1056   | 4  | AY344581 Bos tauru |
| 23 | 652.2 | 43.1 | 1732   | 5  | GGA535754          |
| 24 | 652.2 | 43.1 | 2546   | 5  | AJ720352 Gallus ga |
| 25 | 608.6 | 40.2 | 145414 | 9  | HS392M18           |
| 26 | 586.4 | 38.7 | 1502   | 5  | BC082519           |
| 27 | 586.4 | 38.7 | 1538   | 5  | AJ781498           |
| 28 | 563.6 | 37.2 | 1155   | 5  | AJ606070 Fugu rubr |
| 29 | 552.8 | 36.5 | 1505   | 5  | XLA514425          |
| 30 | 539   | 35.6 | 1164   | 5  | AJ781504           |
| 31 | 538.4 | 35.6 | 1549   | 5  | DRE579536          |
| 32 | 471.4 | 31.1 | 1524   | 9  | BC000582           |
| 33 | 445.4 | 29.4 | 920    | 5  | AJ719585           |
| 34 | 201.6 | 13.3 | 1841   | 3  | AK112708           |
| 35 | 193.4 | 12.8 | 2393   | 9  | AK124093           |
| 36 | 188.8 | 12.5 | 240336 | 2  | AC094844           |
| 37 | 186.6 | 12.3 | 240510 | 2  | AC134371           |
| 38 | 184.6 | 12.2 | 200329 | 10 | AC078911           |
| 39 | 184.6 | 12.2 | 208614 | 10 | AL807380           |
| 40 | 184   | 12.2 | 1146   | 3  | AJ781502           |
| 41 | 172.8 | 11.4 | 1209   | 3  | AJ831490           |
| 42 | 168   | 11.1 | 1295   | 3  | AY118651           |
| 43 | 168   | 11.1 | 1320   | 6  | ARS09393           |
| 44 | 166.8 | 11.0 | 1209   | 3  | AJ781501           |
| 45 | 166.4 | 11.0 | 1209   | 6  | CQ590403           |

ALIGNMENTS

|            |                                     |                                    |         |     |        |                 |
|------------|-------------------------------------|------------------------------------|---------|-----|--------|-----------------|
| RESULT 1   | AR163451                            | Sequence 2 from patent US 6270987. | 1514 bp | DNA | linear | PAT 17-OCT-2001 |
| LOCUS      | AR163451                            |                                    |         |     |        |                 |
| DEFINITION | AR163451                            |                                    |         |     |        |                 |
| ACCESSION  | AR163451                            |                                    |         |     |        |                 |
| VERSION    | AR163451.1                          | GI:16234056                        |         |     |        |                 |
| KEYWORDS   | Unknown.                            |                                    |         |     |        |                 |
| SOURCE     | Unknown.                            |                                    |         |     |        |                 |
| ORGANISM   | Unclassified.                       |                                    |         |     |        |                 |
| REFERENCE  | 1 (bases 1 to 1514)                 |                                    |         |     |        |                 |
| AUTHORS    | Wang, Y. and Spellman, M.W.         |                                    |         |     |        |                 |
| TITLE      | O-fucosyltransferase                |                                    |         |     |        |                 |
| JOURNAL    | Patent: US 6270987-A 2 07-AUG-2001; |                                    |         |     |        |                 |
| FEATURES   | Location/Qualifiers                 |                                    |         |     |        |                 |
| source     | 1..1514                             |                                    |         |     |        |                 |
| ORIGIN     | /organism="unknown"                 |                                    |         |     |        |                 |
|            | /mol_type="unassigned DNA"          |                                    |         |     |        |                 |

|                       |                 |  |           |              |
|-----------------------|-----------------|--|-----------|--------------|
| Query Match           | 100.0%;         | Score 1514;  | DB 6;     | Length 1514; |
| Best Local Similarity | 100.0%;         | Pred. No. 0;   |           |              |
| Matches 1514;         | Conservative 0; | Mismatches 0;  | Indels 0; | Gaps 0;      |
| Qy                    | 1               | ATGCCCCGGGCTCTCTGGGACCCGGCGGTTACCTGTCTACTGCTCCCTGATGGGGCGC   | 60        |              |
| Db                    | 1               | ATGCCCCGGGCTCTCTGGGACCCGGCGGTTACCTGTCTACTGCTCCCTGATGGGGCGC   | 60        |              |
| Qy                    | 61              | TTTGGGAACCAAGCGGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT | 120       |              |
| Db                    | 61              | TTTGGGAACCAAGCGGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT | 120       |              |
| Qy                    | 121             | ACCTTGGCTGCTCCCTCTGGATTGAGTACAGCATCACAGCTCCCTTTCCACCAACCTC   | 180       |              |
| Db                    | 121             | ACCTTGGCTGCTCCCTCTGGATTGAGTACAGCATCACAGCTCCCTTTCCACCAACCTC   | 180       |              |
| Qy                    | 181             | CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTGATC    | 240       |              |
| Db                    | 181             | CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTGATC    | 240       |              |
| Qy                    | 241             | AGCTTGGAGGATTTCTGGAGAACTGGCACCACCACTGGCCCTTGAGAGCGGTG        | 300       |              |
| Db                    | 241             | AGCTTGGAGGATTTCTGGAGAACTGGCACCACCACTGGCCCTTGAGAGCGGTG        | 300       |              |

Db 241 AGCTTGGAGATTTCATGGAGAGCTGGCAACCCACCACTGGCCCCCTTGAGAGCGGGTG 300  
Qy 301 GCATACGCTTTTGGAGTGGCGCCAGCCAGCAAGCCAGATGAAGAAGCTGGCCCATGAAG 360  
Db 301 GCATACGCTTTTGGAGTGGCGCCAGCCAGCAAGCCAGATGAAGAAGCTGGCCCATGAAG 360  
Qy 361 GAAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTTCAATGTAGTTTCAACAAGTCGGAG 420  
Db 361 GAAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTTCAATGTAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTCACAGCATTTTCTTCAAGTCTTCTTACAGAGCAAAATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTTCACAGCATTTTCTTCAAGTCTTCTTACAGAGCAAAATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGGTGTTCCTGCTGACAGAGCCAGCCAGCTTCCCGTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTTCCTGCTGACAGAGCCAGCCAGCTTCCCGTCTAGAA 540  
Qy 541 GAAACAGGCCACTACAGAAATCATGTATGTGTGATGCTGAGCAAGATGTGAAGCGGGAG 600  
Db 541 GAAACAGGCCACTACAGAAATCATGTATGTGTGATGCTGAGCAAGATGTGAAGCGGGAG 600  
Qy 601 GCCAGATTATGCCACCTTTGCGGCCCTATGTGGGCATTCACTGCGCATTTGGCTTCT 660  
Db 601 GCCAGATTATGCCACCTTTGCGGCCCTATGTGGGCATTCACTGCGCATTTGGCTTCT 660  
Qy 661 GACTGGAAGAGCGCTTGCCATGCTGAAGAGACGGGACTGCGAGCTCGCATTTATGCC 720  
Db 661 GACTGGAAGAGCGCTTGCCATGCTGAAGAGACGGGACTGCGAGCTCGCATTTATGCC 720  
Qy 721 TCTCGCAGTGTGTGGCTTACAGCCGACAGCCAGCGGCCCTTCAAGTACTATGTGC 780  
Db 721 TCTCGCAGTGTGTGGCTTACAGCCGACAGCCAGCGGCCCTTCAAGTACTATGTGC 780  
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTGTGGTATGCC 840  
Db 781 CTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTGTGGTATGCC 840  
Qy 841 CAGTGGCTACGTTGCTACTGATTCGAGATTATGTGCTGAGCTTCAACAGCTTCTC 900  
Db 841 CAGTGGCTACGTTGCTACTGATTCGAGATTATGTGCTGAGCTTCAACAGCTTCTC 900  
Qy 901 AAGGGAAGGTGAAGTGTGAGCTTGAAGCTGAGTGGCCAGGCTGACCTGTACATC 960  
Db 901 AAGGGAAGGTGAAGTGTGAGCTTGAAGCTGAGTGGCCAGGCTGACCTGTACATC 960  
Qy 961 CTGCGCAAGCGCACCTTTATTTGGCAACTGTGTCTCTCTCTTCTACTGCTTTGTGAAG 1020  
Db 961 CTGCGCAAGCGCACCTTTATTTGGCAACTGTGTCTCTCTCTTCTACTGCTTTGTGAAG 1020  
Qy 1021 CGGAGCGGGACCTCCAGGGGAGGCGCTTCTTTCTTTCGGCATGGACGGCCCTTAAG 1080  
Db 1021 CGGAGCGGGACCTCCAGGGGAGGCGCTTCTTTCTTTCGGCATGGACGGCCCTTAAG 1080  
Qy 1081 CTGCGGGACGAGTCTGATTTCTGCGGAGACACAGACCTCTGATCTTGGAGGACACAG 1140  
Db 1081 CTGCGGGACGAGTCTGATTTCTGCGGAGACACAGACCTCTGATCTTGGAGGACACAG 1140  
Qy 1141 AGTCTGAGCTGCTTCCAGCCAGGCGCTTGGCAGCCAGAGTGTCTCCGGATTGCAACT 1200  
Db 1141 AGTCTGAGCTGCTTCCAGCCAGGCGCTTGGCAGCCAGAGTGTCTCCGGATTGCAACT 1200  
Qy 1201 CTTCTTCTCAGCTGCAAGATGGAGAGTCCAGGACCCCTTCAAGAGGAGGAGCG 1260  
Db 1201 CTTCTTCTCAGCTGCAAGATGGAGAGTCCAGGACCCCTTCAAGAGGAGGAGCG 1260  
Qy 1261 TCCATATCCAGGGCATAGGACTTTGAGAGTTCCTTAGAGAGAGAGCATCTCCATCGCAC 1320  
Db 1261 TCCATATCCAGGGCATAGGACTTTGAGAGTTCCTTAGAGAGAGAGCATCTCCATCGCAC 1320  
Qy 1321 GTGCTTTCTGCTTCTTCTGGGAATTTCTCACTTGGCAAGAGCTCCAGCTTCGGTCTTCT 1380  
Db 1321 GTGCTTTCTGCTTCTTCTGGGAATTTCTCACTTGGCAAGAGCTCCAGCTTCGGTCTTCT 1380

Qy 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTTTCAGAGAGATTTTTTATAGA 1440  
Db 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTTTCAGAGAGATTTTTTATAGA 1440  
Qy 1441 GAGATTTCTATAAATTTTGATACAAAGTCAATGATCTCTAGAACTCTCTGTGTTTTGA 1500  
Db 1441 GAGATTTCTATAAATTTTGATACAAAGTCAATGATCTCTAGAACTCTCTGTGTTTTGA 1500  
Qy 1501 AAATCATTGAATTC 1514  
Db 1501 AAATCATTGAATTC 1514  
  
RESULT 2  
BD103188  
LOCUS O-fucosyltransferase. 1514 bp DNA linear PAT 27-AUG-2002  
DEFINITION O-fucosyltransferase.  
ACCESSION BD103188.1 GI:22648762  
VERSION BD103188.1  
KEYWORDS JF001527389-A/1.  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1514)  
AUTHORS Wang, Y. and Spellman, M. W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: JP 2001527389-A 1 25-DEC-2001;  
GENENTECH INC  
COMMENT OS Unidentified  
PN JP 2001527389-A/1  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877  
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI  
YANG WANG, MICHAEL W SPELLMAN  
PC C12N15/54, C12N9/10, C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
FH Key Location/Qualifiers  
FT source 1..1514  
FT /organism='Unidentified'.  
  
FEATURES  
source  
1..1514  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
  
ORIGIN  
Query Match 100.0%; Score 1514; DB 6; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ATGCCCGGGCTCTGGGACCCGGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Db 1 ATGCCCGGGCTCTGGGACCCGGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Qy 61 TTTGGGAACAGCGCGGATCACTTCTTGGGCTCTTGGCATTTGCAAAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACAGCGCGGATCACTTCTTGGGCTCTTGGCATTTGCAAAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCTCAAAAGCTCTCTTTCAGCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCTCAAAAGCTCTCTTTCAGCAACCTC 180  
Qy 181 CATGTCCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240  
Db 181 CATGTCCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240  
Qy 241 AGTTGGAGGATTTTCAGAGAGCTGGACACCCACCCCTGSCCCCTGAGAGCGGGTG 300  
Db 241 AGTTGGAGGATTTTCAGAGAGCTGGACACCCACCCCTGSCCCCTGAGAGCGGGTG 300



301 GCATCTGCTTTAGGTGGCAGCCAGCGAAGCCAGATAAGAACAGCTGCCCATGAAG 360  
Db |  
301 GCATCTGCTTTAGGTGGCAGCCAGCGAAGCCAGATAAGAACAGCTGCCCATGAAG 360  
Qy |  
361 GAAGAAACCCCTTTGGCCCATCTTGGGANTCAGTTTCATGTAGTTTCAACAAGTCGGAG 420  
Db |  
361 GAAGAAACCCCTTTGGCCCATCTTGGGANTCAGTTTCATGTAGTTTCAACAAGTCGGAG 420  
Qy |  
421 CTTTTTACAGGCAATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 480  
Db |  
421 CTTTTTACAGGCAATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 480  
Qy |  
481 CAAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCGTCCTAGAA 540  
Db |  
481 CCAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCGTCCTAGAA 540  
Qy |  
541 GAACACAGGCCCATCAGAACTACAGTGTATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAG 600  
Db |  
541 GAACACAGGCCCATCAGAACTACAGTGTATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAG 600  
Qy |  
601 GCCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 660  
Db |  
601 GCCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 660  
Qy |  
661 GACTGGAAGAACCCCTGTGCCATGTGAAGACGGGACTGAGGCTCGCACTTCATGACC 720  
Db |  
661 GACTGGAAGAACCCCTGTGCCATGTGAAGACGGGACTGAGGCTCGCACTTCATGACC 720  
Qy |  
721 TCTCCGAGTGTGGGTACAGCCGACAGCAGCGGCCCTCCATGACATGATGTC 780  
Db |  
721 TCTCCGAGTGTGGGTACAGCCGACAGCAGCGGCCCTCCATGACATGATGTC 780  
Qy |  
781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAGCTCTGGTGAGTCTGGTGGATGCC 840  
Db |  
781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAGCTCTGGTGAGTCTGGTGGATGCC 840  
Qy |  
841 CAGTCGGTCTAGTGTCTACTGATTCGAGAGTATGTGCTCTGAGCTCCACAGCTCTTC 900  
Db |  
841 CAGTCGGTCTAGTGTCTACTGATTCGAGAGTATGTGCTCTGAGCTCCACAGCTCTTC 900  
Qy |  
901 AAAGGAAGGTGAAGGTGTGAGCTGAAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 960  
Db |  
901 AAAGGAAGGTGAAGGTGTGAGCTGAAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 960  
Qy |  
961 CTGCGCAGCGGACCACTTTATTTGGCACTGTGTCTCTCTTCTTCTTCTTCTTCTTCT 1020  
Db |  
961 CTGCGCAGCGGACCACTTTATTTGGCACTGTGTCTCTCTTCTTCTTCTTCTTCTTCT 1020  
Qy |  
1021 CGGAGCGGAGCCTCCAGGGAGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1080  
Db |  
1021 CGGAGCGGAGCCTCCAGGGAGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1080  
Qy |  
1081 CTGCGGAGCAGTCTCTGATTCGCGGAGCACCAGCCCTCTGATCTGAGGAGGAGCCAG 1140  
Db |  
1081 CTGCGGAGCAGTCTCTGATTCGCGGAGCACCAGCCCTCTGATCTGAGGAGGAGCCAG 1140  
Qy |  
1141 AGTCGAGCTGGTCTTCCAGCCAGCTTGGCAGCAGAGTGTCTCCGGGATTCGAACCT 1200  
Db |  
1141 AGTCGAGCTGGTCTTCCAGCCAGCTTGGCAGCAGAGTGTCTCCGGGATTCGAACCT 1200  
Qy |  
1201 CCTCTTCTCAGCTGCAAGATGGAAGAGAGTGCAGGAGCCCTTCAAGGAGGAGAGCGC 1260  
Db |  
1201 CCTCTTCTCAGCTGCAAGATGGAAGAGAGTGCAGGAGCCCTTCAAGGAGGAGAGCGC 1260  
Qy |  
1261 TCCATATCCAGGGGATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
Db |  
1261 TCCATATCCAGGGGATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
Qy |  
1321 GTGCTTCTGCTTCTTCTGGAATTTCTCACACTGCAAGCAGTCCAGCTCTCTTCT 1380  
Db |  
1321 GTGCTTCTGCTTCTTCTGGAATTTCTCACACTGCAAGCAGTCCAGCTCTCTTCT 1380  
Qy |  
1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGMACTCTTCTCAGAGAGATTTTTTTATAGA 1440

1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGMACTCTTCTCAGAGAGATTTTTTTATAGA 1440  
Qy |  
1441 GAGATTTCTATAATTTTGTATACAAAGTTCATGACTATCTAGAACTCTCTGTGTTTTTGA 1500  
Db |  
1441 GAGATTTCTATAATTTTGTATACAAAGTTCATGACTATCTAGAACTCTCTGTGTTTTTGA 1500  
Qy |  
1501 AAATCATTTGAATTC 1514  
Db |  
1501 AAATCATTTGAATTC 1514  
RESULT 3  
AF375884  
LOCUS  
DEFINITION Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete  
ACCESSION AF375884  
VERSION AF375884.1 GI:15825113  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 5249)  
AUTHORS Wang, Y., Shao, L., Shi, S., Harris, R.J., Spellman, M.W., Stanley, P. and Haltiwanger, R.S.  
TITLE Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase  
J. Biol. Chem. 276 (43), 40338-40345 (2001)  
JOURNAL 21523965  
MEDLINE 11524432  
REFERENCE 2 (bases 1 to 5249)  
AUTHORS Shi, S., Stanley, P., Wang, Y., Shao, L., Harris, R.J., Spellman, M.W. and Haltiwanger, R.S.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave., New York, NY 10461, USA  
FEATURES  
source  
1..5249  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="between PLAGL2 and KIF3B"  
/tissue\_type="heart"  
1..5249  
/gene="POFUT1"  
50..1216  
/gene="POFUT1"  
/functions="catalyzes transfer of fucose from GDP-fucose to serine or threonine on epidermal growth factor-like repeats"  
/note="O-FucT-1; glycosyltransferase"  
/codon\_start=1  
/product="protein o-fucosyltransferase"  
/protein\_id="AA09576.1"  
/db\_xref="GI:15825114"  
ORIGIN  
Query Match 99.9%; Score 1512.4; DB 9; Length 5249;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCGCGGCTCTCTGGGACCGCGGTTACCTGCTCTACTGCTCTACTGCCCTCGATGGCGGC 60

Db 119 ATGCCCGGGGCTCTCTGGAGCCGGCCGGTTACTGCTCTACTGCCCCCTGATCGGGCGC 178  
Qy 61 TTTGGGAACAGGCGCATCACTTCTTTGGGCTCTCTGGCATTTTGCAAAAGCTCTAAACCGT 120  
Db 179 TTTGGGAACAGGCGCATCACTTCTTTGGGCTCTCTGGCATTTTGCAAAAGCTCTAAACCGT 238  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACAGCATCAAAAGCTCTCTTTCAACAACTC 180  
Db 239 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACAGCATCAAAAGCTCTCTTTCAACAACTC 298  
Qy 181 CATGTGCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 299 CATGTGCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 358  
Qy 241 AGCTTGGAGATTTCAATGGAGCTGGACCCACACCTGCGCCCTCGAGAGCGGGT 300  
Db 359 AGCTTGGAGATTTCAATGGAGCTGGACCCACACCTGCGCCCTCGAGAGCGGGT 418  
Qy 301 GATATCTGCTTGAAGTGGAGCCAGCAGAGCCAGATTAAGAAAGCTGCCCATAGAG 360  
Db 419 GATATCTGCTTGAAGTGGAGCCAGCAGAGCCAGATTAAGAAAGCTGCCCATAGAG 478  
Qy 361 GAAGAAACCCCTTTGGCCCATCTGGGATCAGTTTCAATGTAGTTTCAACAAAGTCGAG 420  
Db 479 GAAGAAACCCCTTTGGCCCATCTGGGATCAGTTTCAATGTAGTTTCAACAAAGTCGAG 538  
Qy 421 CTTTTTACAGGCATTTCTTTAGTGTCTTACAGAGAAACAATGGAGCCAGAGATTTTCT 480  
Db 539 CTTTTTACAGGCATTTCTTTAGTGTCTTACAGAGAAACAATGGAGCCAGAGATTTTCT 598  
Qy 481 CCAAGGAACATCCGGTGTCTTGGCTGCGAGAGCCCGAGCTTCCCGCTCTAGAA 540  
Db 599 CCAAGGAACATCCGGTGTCTTGGCTGCGAGAGCCCGAGCTTCCCGCTCTAGAG 658  
Qy 541 GAACACAGGCACCTACAGAGTACATGTTATGTTGTCAGACGAAATGTTGAAGACGGAGAG 600  
Db 659 GAACACAGGCACCTACAGAGTACATGTTATGTTGTCAGACGAAATGTTGAAGACGGAGAG 718  
Qy 601 GCCAGATTCATGCCACCTTGTCCGCTTATGTTGGGATTCATCTGGCATTCGGCTCT 660  
Db 719 GCCAGATTCATGCCACCTTGTCCGCTTATGTTGGGATTCATCTGGCATTCGGCTCT 778  
Qy 661 GACTGGAAGAACGCTGTGCTGATGCTGAAGAGACGGGACTGCGAGCTCGCATTCATGCC 720  
Db 779 GACTGGAAGAACGCTGTGCTGATGCTGAAGAGACGGGACTGCGAGCTCGCATTCATGCC 838  
Qy 721 TCTCCGCAAGTGTGGGCTACAGCCGACACAGCGGCCCGCTTACAGATGATATGTC 780  
Db 839 TCTCCGCAAGTGTGGGCTACAGCCGACACAGCGGCCCGCTTACAGATGATATGTC 898  
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGCTGAGGTCGCTGGATGCC 840  
Db 899 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGCTGAGGTCGCTGGATGCC 958  
Qy 841 CAGTGGTCTAGTGTCTTACTGATTCGAGAGTTATGCTGAGCTTCCACAGCTCTTC 900  
Db 959 CAGTGGTCTAGTGTCTTACTGATTCGAGAGTTATGCTGAGCTTCCACAGCTCTTC 1018  
Qy 901 AAAGGAGAGTGAAGGTGTGAGCTTGAAGCTGAGGTCGAGGTCGAGCTTACATC 960  
Db 1019 AAAGGAGAGTGAAGGTGTGAGCTTGAAGCTGAGGTCGAGGTCGAGCTTACATC 1078  
Qy 961 CTGGCCAAAGCCGACCTATTATGGCACTGTGCTCTCTTCACTGCTTGTGAG 1020  
Db 1079 CTGGCCAAAGCCGACCTATTATGGCACTGTGCTCTCTTCACTGCTTGTGAG 1138  
Qy 1021 CGGAGCGGACCTCCAGGGAGGCGCTTCTTTCTTGGCATGAGACGAGCCCTTAAG 1080  
Db 1139 CGGAGCGGACCTCCAGGGAGGCGCTTCTTTCTTGGCATGAGACGAGCCCTTAAG 1198  
Qy 1081 CTGCGGGACGAGTTCTGATTTGCGCGGACACAGACCTCTGATCTCGAGGGACAG 1140

Db 1199 CTGCGGAGCAGATTCTGATTCTGGCCGGAGCACAGACCCTCTGATCCTGGAGGACCAG 1258  
Qy 1141 AGTCTGAGCTGGTCTCTTCAGCCAGGCTTGGAGGCTCTCCGGATTCGCAAACT 1200  
Db 1259 AGTCTGAGCTGGTCTCTTCAGCCAGGCTTGGAGGCTCTCCGGATTCGCAAACT 1318  
Qy 1201 CCTCTTCTACCTGCCAAAGATGGAGAGAGTGCCAGGACCCCTCAAGAGGGAGAGCG 1260  
Db 1319 CCTCTTCTACCTGCCAAAGATGGAGAGAGTGCCAGGACCCCTCAAGAGGGAGAGCG 1378  
Qy 1261 TCCATATCCAGGAGCATAGGACTTTCAGGCTTCTAGGAGCAGAGCATCTCCCATCGCAC 1320  
Db 1379 TCCATATCCAGGAGCATAGGACTTTCAGGCTTCTAGGAGCAGAGCATCTCCCATCGCAC 1438  
Qy 1321 GTGCTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGGCTCCGCTTCT 1380  
Db 1439 GTGCTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGGCTCCGCTTCT 1498  
Qy 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAAGCTTTTCAGAGAGATTTTATAGA 1440  
Db 1499 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAAGCTTTTCAGAGAGATTTTATAGA 1558  
Qy 1441 GAGATTTCTATATTTTGTATACAGGTGATGCTGAAGCTTTTCAGAGAGATTTTATAGA 1500  
Db 1559 GAGATTTCTATATTTTGTATACAGGTGATGCTGAAGCTTTTCAGAGATTTTATAGA 1618  
Qy 1501 AAATCATTTGAATTC 1514  
Db 1619 AAATCATTTGAATTC 1632

RESULT 4  
BD103190  
LOCUS BD103190  
DEFINITION O-fucosyltransferase.  
ACCESSION BD103190  
VERSION BD103190.1 GI:22648764  
KEYWORDS JP 2001527389-A/3.  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 11284)  
AUTHORS Wang, Y. and Spellman, M. W.  
TITLE O-fucosyltransferase.  
JOURNAL Patent JP 2001527389-A 3 25-DEC-2001;  
COMMENT GENENTECH INC  
OS Unidentified  
PN JP 2001527389-A/3  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877  
PR 31-JAN-1997 US 08/792498,26-NOV-1997 US 08/978741 PI  
PC C12N15/54,C12N9/10,C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
FH Key Location/Qualifiers  
FT source 1..11284  
FT /organism='Unidentified'.  
FEATURES  
source Location/Qualifiers  
1..11284  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'

ORIGIN  
Query Match 99.9%; Score 1512.4; DB 6; Length 11284;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCGGGGCTCTCTGGAGCCGGCCGGTTACTGCTCTACTGCCCCCTGATCGGGCGC 60  
Db 4236 ATGCCCGGGGCTCTCTGGAGCCGGCCGGTTACTGCTCTACTGCCCCCTGATCGGGCGC 4295

QY 61 TTTGGGAACCCAGGCGGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 120  
Db |||||  
4296 TTTGGGAACCCAGGCGGATCACTTCTTGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 4355  
QY 121 ACCTTGGCTGTCCTTCTTGGATTAAGTACAGAGATCAAGAGCTCTCTTACCAACCTC 180  
Db |||||  
4356 ACCTTGGCTGTCCTTCTTGGATTAAGTACAGAGATCAAGAGCTCTCTTACCAACCTC 4415  
QY 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240  
Db |||||  
4416 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 4475  
QY 241 AGCTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCTCGGCCCTTGAGAAGCGGGTG 300  
Db |||||  
4476 AGCTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCTCGGCCCTTGAGAAGCGGGTG 4535  
QY 301 GCATACCTCTTGGAGTGGAGCCAGGAGAGCCAGATAGAGAGCTGCCCATGAAG 360  
Db |||||  
4536 GCATACCTCTTGGAGTGGAGCCAGGAGAGCCAGATAGAGAGCTGCCCATGAAG 4595  
QY 361 GAAGAAACCCCTTTGGGCCATCTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db |||||  
4596 GAAGAAACCCCTTTGGGCCATCTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 4655  
QY 421 CTTTTTACAGGCATTTCTTCAAGTCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 480  
Db |||||  
4656 CTTTTTACAGGCATTTCTTCAAGTCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 4715  
QY 481 CCAAGGAACATCGGTCCTTGCCTGCGAGAGCCCGAGCCAGTTCCTCCGTCCTAGAA 540  
Db |||||  
4716 CCAAGGAACATCGGTCCTTGCCTGCGAGAGCCCGAGCCAGTTCCTCCGTCCTAGAG 4775  
QY 541 GAAACACGCGCCATACAGAGTACATGTATGTGTCAGACGAAATGGTGAACACGGGAGAG 600  
Db |||||  
4776 GAAACACGCGCCATACAGAGTACATGTATGTGTCAGACGAAATGGTGAACACGGGAGAG 4835  
QY 601 GCCCAGATTATGCCCATCTTGTCCGGCCCTATGTGGCCATTCATCTGCGCATTTGGTCT 660  
Db |||||  
4836 GCCCAGATTATGCCCATCTTGTCCGGCCCTATGTGGCCATTCATCTGCGCATTTGGTCT 4895  
QY 661 GACTGGAAGACCGCTGTGCGCATGTGAAGACGGGATGTGAGGCTGCGCATTTGATGCC 720  
Db |||||  
4896 GACTGGAAGACCGCTGTGCGCATGTGAAGACGGGATGTGAGGCTGCGCATTTGATGCC 4955  
QY 721 TCTCGCAGTGTGGGCTACAGCGCAGCAGCGGCCCTCCAGCATCTATGTGC 780  
Db |||||  
4956 TCTCGCAGTGTGGGCTACAGCGCAGCAGCGGCCCTCCAGCATCTATGTGC 5015  
QY 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGCC 840  
Db |||||  
5016 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGCC 5075  
QY 841 CAGTCGCTACTAGTTCTGATTTCCAGAGTTATGTGCTCTGAGCTCCAAAGCTCTTC 900  
Db |||||  
5076 CAGTCGCTACTAGTTCTGATTTCCAGAGTTATGTGCTCTGAGCTCCAAAGCTCTTC 5135  
QY 901 AAAGGAAGGTGAAGTGTGAGCTGAAGCTGAGGTGGCCAGGTCGACCTGTACATC 960  
Db |||||  
5136 AAAGGAAGGTGAAGTGTGAGCTGAAGCTGAGGTGGCCAGGTCGACCTGTACATC 5195  
QY 961 CTCGCCCAAGCCGACCATTTTATTGGCACTGTGTCTCTCTTCACTGCTCTTGTGAAG 1020  
Db |||||  
5196 CTCGCCCAAGCCGACCATTTTATTGGCACTGTGTCTCTCTTCACTGCTCTTGTGAAG 5255  
QY 1021 CGGAGCGGACCTTCCAGGGAGGCGGCTCTTCTTTCGGGATGGACAGGCCCTTAAAG 1080  
Db |||||  
5256 CGGAGCGGACCTTCCAGGGAGGCGGCTCTTCTTTCGGGATGGACAGGCCCTTAAAG 5315  
QY 1081 CTGGGGACGAGTCTGATTTCTGCGGAGCAGCAGACCTCTGATCTGAGGGAGCCAG 1140  
Db |||||  
5316 CTGGGGACGAGTCTGATTTCTGCGGAGCAGCAGACCTCTGATCTGAGGGAGCCAG 5375

QY 1141 AGTCTGAGCTGCTCTTCCAGCCAGGCTGCGACGAGGTGCTCCGGGATTTGCAAACT 1200  
Db |||||  
5376 AGTCTGAGCTGCTCTTCCAGCCAGGCTGCGACGAGGTGCTCCGGGATTTGCAAACT 5435  
QY 1201 CCTCTTCTCACTGCCCAAGATGGAGAAGTGGAGGACCCCTCAAGAGGGAGAGCGC 1260  
Db |||||  
5436 CCTCTTCTCACTGCCCAAGATGGAGAAGTGGAGGACCCCTCAAGAGGGAGAGCGC 5495  
QY 1261 TCCATATCCAGGCGATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
Db |||||  
5496 TCCATATCCAGGCGATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 5555  
QY 1321 GTGCTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTTCT 1380  
Db |||||  
5556 GTGCTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTTCT 5615  
QY 1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATTTTATATAGA 1440  
Db |||||  
5616 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATTTTATATAGA 5675  
QY 1441 GAGATTTCTATAATTTTGATACAAAGGTCTGACTATCTAGAACTCTCTGTGGTTTTTGA 1500  
Db |||||  
5676 GAGATTTCTATAATTTTGATACAAAGGTCTGACTATCTAGAACTCTCTGTGGTTTTTGA 5735  
QY 1501 AAATCATTTGAATTC 1514  
Db |||||  
5736 AAATCATTTGAATTC 5749

RESULT 5  
LOCUS D80002 5189 bp mRNA linear PRI 10-JAN-2003  
DEFINITION Homo sapiens KIAA0180 mRNA, complete cds.  
ACCESSION D80002  
VERSION D80002.2 GI:20521837  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS Nagase, T., Seki, N., Iehikawa, K., Tanaka, A. and Nomura, N.  
TITLE Prediction of the coding sequences of unidentified human genes. V.  
The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by  
analysis of cDNA clones from human cell line KG-1  
JOURNAL DNA Res. 3 (1), 17-24 (1996)  
MEDLINE 96281124  
REFERENCE 8724849  
2  
AUTHORS Chiang, P.W., Wang, S., Smithivas, P., Song, W.J., Ramamoorthy, S.,  
Hillman, J., Puett, S., Van Keuren, M.L., Crombez, E., Kumar, A.,  
Glover, T.W., Miller, D.E., Tsai, C.H., Blackburn, C.C., Chen, X.N.,  
Sun, Z., Cheng, J.F., Korenberg, J.R. and Kurnit, D.M.  
TITLE Identification and analysis of the human and murine putative  
chromatin structure regulator SUPT6H and Supt6h  
JOURNAL Genomics 34 (3), 328-333 (1996)  
MEDLINE 96374824  
REFERENCE 8786132  
3 (bases 1 to 5189)  
AUTHORS Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;  
1532-3, Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)  
COMMENT On May 9, 2002 this sequence version replaced gi:1136419.  
FEATURES  
source Location/Qualifiers  
1. 5189  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="ha0256781"  
/sex="male"  
/cell\_line="KG-1"



AUTHORS Haerlach, T., Schoch, C., Kern, W., Kohlmann, A., Schnittger, S.,  
Dugas, M., Eils, R., Bror, B. and Mergenthaler, S.  
TITLE Novel genetic markers for leukemias  
JOURNAL Patent: WO 03039443-A 2310 15-MAY-2003;  
Deutsches Krebsforschungszentrum (DE);  
Ludwig-Maximilians-Universitaet Muenchen (DE); Haerlach, Torsten,  
PD Dr. Dr. Schoch, Claudia (DE); Kern, Wolfgang (DE)  
FEATURES  
source Location/Qualifiers  
1..5266  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 99.7%; Score 1509.8; DB 6; Length 5266;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 151; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGCCGCGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCAATGGGGCGC 60  
DB 132 ATGCCGCGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCAATGGGGCGC 191  
QY 61 TTTGGGAACCAAGCGGATCACTTCTTGGGCTCTGGGCAATTCGAAAGCTGCTAAACCGT 120  
DB 192 TTTGGGAACCAAGCGGATCACTTCTTGGGCTCTGGGCAATTCGAAAGCTGCTAAACCGT 251  
QY 121 ACCTTGGCTGTCCTCTTGGGATTCAGTACAGATCAAGAGCTCTTTCACCAACCTC 180  
DB 252 ACCTTGGCTGTCCTCTTGGGATTCAGTACAGATCAAGAGCTCTTTCACCAACCTC 311  
QY 181 CATGTGTCCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240  
DB 312 CATGTGTCCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 371  
QY 241 AGCTTGGAGGATTCATGGAGAGCTGCGACCCACCCAGCTGCCCTCCAGAGCGGGTG 300  
DB 372 AGCTTGGAGGATTCATGGAGAGCTGCGACCCACCCAGCTGCCCTCCAGAGCGGGTG 431  
QY 301 GCATACCTCTTGGAGTGGAGCCAGCGAAGCCAGATAGAGAGCTGCCCATGAAG 360  
DB 432 GCATACCTCTTGGAGTGGAGCCAGCGAAGCCAGATAGAGAGCTGCCCATGAAG 491  
QY 361 GAAGAAACCCCTTTGGCCCATTCCTGGGATCAGTTTCATGTAGTTTCAACAAGTCGGAG 420  
DB 492 GAAGAAACCCCTTTGGCCCATTCCTGGGATCAGTTTCATGTAGTTTCAACAAGTCGGAG 551  
QY 421 CTTTITACAGGATTCCTTCAGTGCTTCCTACAGAGAAATGAGGACGAGATTTTCT 480  
DB 552 CTTTITACAGGATTCCTTCAGTGCTTCCTACAGAGAAATGAGGACGAGATTTTCT 611  
QY 481 CCAAAGGAACATCCGGTCTTCCCTGCGAGGAGCCCGAGCTTCCCGCTCTAGAA 540  
DB 612 CCAAAGGAACATCCGGTCTTCCCTGCGAGGAGCCCGAGCTTCCCGCTCTAGAG 671  
QY 541 GAACACAGGCCATACAGAAGTACATGTATGGTTCAGAGAAATGGTGAAGACGGGAGAG 600  
DB 672 GAACACAGGCCATACAGAAGTACATGTATGGTTCAGAGAAATGGTGAAGACGGGAGAG 731  
QY 601 GCCCAGATTCATGCCACCTTGTCCGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 660  
DB 732 GCCCAGATTCATGCCACCTTGTCCGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 791  
QY 661 GACTGGAAGACGCTGTGCCATGCTGAAGACGGGAGCTGAGGCTGCACTTCATGGCC 720  
DB 792 GACTGGAAGACGCTGTGCCATGCTGAAGACGGGAGCTGAGGCTGCACTTCATGGCC 851  
QY 721 TCTCCGCAAGTGTGGGCTACAGCGGACAGCGGCCCCCTCCAGCATGACTATGTGC 780  
DB 852 TCTCCGCAAGTGTGGGCTACAGCGGACAGCGGCCCCCTCCAGCATGACTATGTGC 911  
QY 781 CTGCTCAGCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 840  
DB 912 CTGCTCAGCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 971

QY 841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900  
DB 972 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 1031  
QY 901 AAAGGAAGGTGAAGGTGGTGAAGCTGAGGCTGAGGTGGCCAGGTCGACCTGTACATC 960  
DB 1032 AAAGGAAGGTGAAGGTGGTGAAGCTGAGGCTGAGGTGGCCAGGTCGACCTGTACATC 1091  
QY 961 CTGCGCAAGCGGACCACTTTATTGGCAACTGTGCTCTCTCTTACACCTCTTGTGAAG 1020  
DB 1092 CTGCGCAAGCGGACCACTTTATTGGCAACTGTGCTCTCTCTTACACCTCTTGTGAAG 1151  
QY 1021 CCGGAGCGGACCTCCAGGGGAGGCGCTCTTTCTTCTCGCATGGACAGGCGCCCTAAG 1080  
DB 1152 CCGGAGCGGACCTCCAGGGGAGGCGCTCTTTCTTCTCGCATGGACAGGCGCCCTAAG 1211  
QY 1081 CTGCGGACGAGTCTGATTCTGGCCGGAGCACAGACCTCTGATCTGAGGAGGACCGAG 1140  
DB 1212 CTGCGGACGAGTCTGATTCTGGCCGGAGCACAGACCTCTGATCTGAGGAGGACCGAG 1271  
QY 1141 AGTCTGAGCTGGTCTCTTCCAGCCAGGCTGCGAGCAGAGTGTCTCCGGGATTCGAACT 1200  
DB 1272 AGTCTGAGCTGGTCTCTTCCAGCCAGGCTGCGAGCAGAGTGTCTCCGGGATTCGAACT 1331  
QY 1201 CCTCTTCTCAGCTGCCAAAGATGAGAGAGTGCAGGAGCCCTCAAGGAGGAGAGCGC 1260  
DB 1332 CCTCTTCTCAGCTGCCAAAGATGAGAGAGTGCAGGAGCCCTCAAGGAGGAGAGCGC 1391  
QY 1261 TCCATATCCAGGCGATAGGACTTTGCAGGTTCTTAGGAGGAGGAGCATCTCCCATCGCAC 1320  
DB 1392 TCCATATCCAGGCGATAGGACTTTGCAGGTTCTTAGGAGGAGGAGCATCTCCCATCGCAC 1451  
QY 1321 GTGCTTCTGCTCTTCTGGGAAATTTCTCAGACTGGCAAGAGCTCCAGCTCTCTCTTCT 1380  
DB 1452 GTGCTTCTGCTCTTCTGGGAAATTTCTCAGACTGGCAAGAGCTCCAGCTCTCTCTTCT 1511  
QY 1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATTTTTTATAGA 1440  
DB 1512 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATTTTTTATAGA 1571  
QY 1441 GAGATTTCTATAATTTTGTATACAGGTCATGACTATCTAGAACTCTCTGTGTTTTGA 1500  
DB 1572 GAGATTTCTATAATTTTGTATACAGGTCATGACTATCTAGAACTCTCTGTGTTTTGA 1631  
QY 1501 AAATCATTGAATTC 1514  
DB 1632 AAATCATTGAATTC 1645

RESULT 7  
HSM803330  
LOCUS HSM803330 4560 bp mRNA linear PRI 13-MAY-2003  
DEFINITION Homo sapiens mRNA; cDNA DKF2p451j114 (from clone DKF2p451j114).  
ACCESSION AL832023  
VERSION AL832023.1 GI:21732563  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 4560)  
AUTHORS Wambutt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,  
Fobo, G., Han, M. and Wiemann, S.  
DIRECT SUBMISSION  
TITLE Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764  
JOURNAL Neuherberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing  
consortium of the German Genome Project.  
This clone (DKF2p451j114) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059



| Qy | Db | 1021  | 1146  | 1081  | 1206  | 1141  | 1266  | 1201  | 1326   | 1261   | 1386  | 1321  | 1446  | 1381  | 1506  | 1441                                   |
|----|----|---|---|---|---|---|---|---|--|--|---|---|---|---|---|--|
|    |    |   |   |   |   |   |   |   |  |  |   |   |   |   |   |  |
|    |    | GGACGAGTTCTGATTTCTGGCGGAGCACCAGACCCCTCTGATCTCTGGAGGACCAGAGTCT | GAGCTGTCCTTCCAGCCAGSCCTGGCAGCCAGAGTGCTCCGGGATTGCAAACTCCCTCT | GAGCTGTCCTTCCAGCCAGSCCTGGCAGCCAGAGTGCTCCGGGATTGCAAACTCCCTCT | TCTCACCTGTCACCAAGATGGAGAAGAGTGCCAGAGGACCCCTCAAGAGGGAGAGCGCTCCAT | TCTCACCTGTCACCAAGATGGAGAAGAGTGCCAGAGGACCCCTCAAGAGGGAGAGCGCTCCAT | ATCCAGGGCATAGGACTTTCAGAGTTCCTTAGAGCAGGAGCATCTCCATCGCACGTGCT | ATCCAGGGCATAGGACTTTCAGAGTTCCTTAGAGCAGGAGCATCTCCATCGCACGTGCT | ATCCAGGGCATAGGACTTTCAGAGTTCCTTAGAGCAGGAGCATCTCCATCGCACGTGCT  | TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCCTCCGTTCTTCTGGTCC | TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCCTCCGTTCTTCTGGTCC  | ACTCTGCTGTGACGACGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTTATAGAGAGAT | ACTCTGCTGTGACGACGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTTATAGAGAGAT | TTCTATAATTTTGATACAGGTCATGACTATCTAGAACTCTCTGTGGTTTTTTTGAATAATC | TTCTATAATTTTGATACAGGTCATGACTATCTAGAACTCTCTGTGGTTTTTTTGAATAATC | ATTGAATTC 1514                         |
|    |    | 1021  | 1146  | 1081  | 1206  | 1141  | 1266  | 1201  | 1326   | 1261   | 1386  | 1321  | 1446  | 1381  | 1506  | 1441                                   |
|    |    |   |   |   |   |   |   |   |  |  |   |   |   |   |   |  |
|    |    | GGACGAGTTCTGATTTCTGGCGGAGCACCAGACCCCTCTGATCTCTGGAGGACCAGAGTCT | GAGCTGTCCTTCCAGCCAGSCCTGGCAGCCAGAGTGCTCCGGGATTGCAAACTCCCTCT | GAGCTGTCCTTCCAGCCAGSCCTGGCAGCCAGAGTGCTCCGGGATTGCAAACTCCCTCT | TCTCACCTGTCACCAAGATGGAGAAGAGTGCCAGAGGACCCCTCAAGAGGGAGAGCGCTCCAT | TCTCACCTGTCACCAAGATGGAGAAGAGTGCCAGAGGACCCCTCAAGAGGGAGAGCGCTCCAT | ATCCAGGGCATAGGACTTTCAGAGTTCCTTAGAGCAGGAGCATCTCCATCGCACGTGCT | ATCCAGGGCATAGGACTTTCAGAGTTCCTTAGAGCAGGAGCATCTCCATCGCACGTGCT | TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCCTCCGTTCTTCTGGTCC | TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCCTCCGTTCTTCTGGTCC | ACTCTGCTGTGACGACGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTTATAGAGAGAT | ACTCTGCTGTGACGACGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTTATAGAGAGAT | TTCTATAATTTTGATACAGGTCATGACTATCTAGAACTCTCTGTGGTTTTTTTGAATAATC | TTCTATAATTTTGATACAGGTCATGACTATCTAGAACTCTCTGTGGTTTTTTTGAATAATC | ATTGAATTC 1514  | ATTGAATTC 1449                         |
|    |    | LOCUS   | BD103191  | BD103191  | BD103191  | BD103191  | BD103191  | BD103191  | BD103191   | BD103191   | BD103191  | BD103191  | BD103191  | BD103191  | BD103191  | BD103191                               |
|    |    | DEFINITION  | O-fucosyltransferase.                                       | O-fucosyltransferase.                                       | O-fucosyltransferase.   | O-fucosyltransferase.   | O-fucosyltransferase.                                       | O-fucosyltransferase.                                       | O-fucosyltransferase.  | O-fucosyltransferase.  | O-fucosyltransferase.   | O-fucosyltransferase.   | O-fucosyltransferase.   | O-fucosyltransferase.   | O-fucosyltransferase.   | O-fucosyltransferase.                  |
|    |    | ACCESSION   | BD103191  | BD103191  | BD103191  | BD103191  | BD103191  | BD103191  | BD103191   | BD103191   | BD103191  | BD103191  | BD103191  | BD103191  | BD103191  | BD103191                               |
|    |    | VERSION   | BD103191.1  | BD103191.1  | BD103191.1  | BD103191.1  | BD103191.1  | BD103191.1  | BD103191.1   | BD103191.1   | BD103191.1  | BD103191.1  | BD103191.1  | BD103191.1  | BD103191.1  | BD103191.1                             |
|    |    | KEYWORDS  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.   | JP 2001527389-A/4.   | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.  | JP 2001527389-A/4.                     |
|    |    | SOURCE  | unidentified  | unidentified  | unidentified  | unidentified  | unidentified  | unidentified  | unidentified   | unidentified   | unidentified  | unidentified  | unidentified  | unidentified  | unidentified  | unidentified                           |
|    |    | ORGANISM  | unclassified.   | unclassified.   | unclassified.   | unclassified.   | unclassified.   | unclassified.   | unclassified.  | unclassified.  | unclassified.   | unclassified.   | unclassified.   | unclassified.   | unclassified.   | unclassified.                          |
|    |    | REFERENCE   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)  | 1 (bases 1 to 5009)  | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)   | 1 (bases 1 to 5009)                    |
|    |    | AUTHORS   | Wang, Y. and Spellman, M.W.                                 | Wang, Y. and Spellman, M.W.                                 | Wang, Y. and Spellman, M.W.                                     | Wang, Y. and Spellman, M.W.                                     | Wang, Y. and Spellman, M.W.                                 | Wang, Y. and Spellman, M.W.                                 | Wang, Y. and Spellman, M.W.                                  | Wang, Y. and Spellman, M.W.                                  | Wang, Y. and Spellman, M.W.                                   | Wang, Y. and Spellman, M.W.                                   | Wang, Y. and Spellman, M.W.                                   | Wang, Y. and Spellman, M.W.                                   | Wang, Y. and Spellman, M.W.                                   | Wang, Y. and Spellman, M.W.            |
|    |    | TITLE   | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase   | O-fucosyltransferase   | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase  | O-fucosyltransferase                   |
|    |    | JOURNAL   | Patent: JP 2001527389-A 4 25-DEC-2001;                      | Patent: JP 2001527389-A 4 25-DEC-2001;                      | Patent: JP 2001527389-A 4 25-DEC-2001;                          | Patent: JP 2001527389-A 4 25-DEC-2001;                          | Patent: JP 2001527389-A 4 25-DEC-2001;                      | Patent: JP 2001527389-A 4 25-DEC-2001;                      | Patent: JP 2001527389-A 4 25-DEC-2001;                       | Patent: JP 2001527389-A 4 25-DEC-2001;                       | Patent: JP 2001527389-A 4 25-DEC-2001;                        | Patent: JP 2001527389-A 4 25-DEC-2001;                        | Patent: JP 2001527389-A 4 25-DEC-2001;                        | Patent: JP 2001527389-A 4 25-DEC-2001;                        | Patent: JP 2001527389-A 4 25-DEC-2001;                        | Patent: JP 2001527389-A 4 25-DEC-2001; |
|    |    | COMMENT   | GENENTECH INC   | GENENTECH   |   |   |   |   |  |  |   |   |   |   |   |  |



Db 1 GAACAGGCGGATCACTCTCTGGGCTCTCTGGCATTTTGCAGGCTGCTAAACCGTACCTT 60  
Qy 126 GGCTCTCCCTCTTGGATTGAGTACAGCATCAAAAGCTCTCTTTTCAACAACTTCCATGT 185  
Db 61 GGCTCTCCCTCTTGGATTGAGTACAGCATCAAAAGCTCTCTTTTCAACAACTTCCATGT 120  
Qy 186 GTCTTACAGAACTACTTCAAGCTGAGCCCTCCAGGCTTACCATCGGGTCAATCAGCTT 245  
Db 121 GTCTTACAGAACTACTTCAAGCTGAGCCCTCCAGGCTTACCATCGGGTCAATCAGCTT 180  
Qy 246 GGAGGATTTTATGGAGAGCTGGCAACCACTGAGCCCTTGAAGAGCGGGTGGGATA 305  
Db 181 GGAGGATTTTATGGAGAGCTGGCAACCACTGAGCCCTTGAAGAGCGGGTGGGATA 240  
Qy 306 CTGCTTTGAGTGGCAGCCAGGAAAGCCAGATAAGAGACGTGCCCCATGAAGGAAGG 365  
Db 241 CTGCTTTGAGTGGCAGCCAGGAAAGCCAGATAAGAGACGTGCCCCATGAAGGAAGG 300  
Qy 366 AAACCCCTTTGGCCCAATCTGGGATCAGTTTCAATGTGAGTTTCAACAAGTGGAGCTTTT 425  
Db 301 AAACCCCTTTGGCCCAATCTGGGATCAGTTTCAATGTGAGTTTCAACAAGTGGAGCTTTT 360  
Qy 426 TACAGGCATTTTCTTCAAGTCTTCTTACAGAGAAACAATGGAGCCAGAGATTTTCTCCAAA 485  
Db 361 TACAGGCATTTTCTTCAAGTCTTCTTACAGAGAAACAATGGAGCCAGAGATTTTCTCCAAA 420  
Qy 486 GGAAACATCCGGTGTCTGCGCTGCGAGGAGCCCGAGCCAGTTCCCGCTCTTGAAGAAACA 545  
Db 421 GGAAACATCCGGTGTCTGCGCTGCGAGGAGCCCGAGCCAGTTCCCGCTCTTGAAGAAACA 480  
Qy 546 CAGGCCACTACAGAACTACATGTTATGTTGATGAGTGAAGAAATGTTGAAGAGAGGCCCA 605  
Db 481 CAGGCCACTACAGAACTACATGTTATGTTGATGAGTGAAGAAATGTTGAAGAGAGGCCCA 540  
Qy 606 GATTTCATGCGCATCTGTTGCGGCGCTATGTTGGGCAATTCATCTCGGATGTTGCTGACTG 665  
Db 541 GATTTCATGCGCATCTGTTGCGGCGCTATGTTGGGCAATTCATCTCGGATGTTGCTGACTG 600  
Qy 666 GAAGAACCCCTTGCCTATGCTGAGAGAGCGGACTGAGGCTGCCATTCATGGGCTCTTCC 725  
Db 601 GAAGAACCCCTTGCCTATGCTGAGAGAGCGGACTGAGGCTGCCATTCATGGGCTCTTCC 660  
Qy 726 GCAGTGTGTGGCTTACAGCGCGAGCAACAGCGGCGCCCTCACGATGATGTGCTGCC 785  
Db 661 GCAGTGTGTGGCTTACAGCGCGAGCAACAGCGGCGCCCTCACGATGATGTGCTGCC 720  
Qy 786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGATGCCAGTC 845  
Db 721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGATGCCAGTC 780  
Qy 846 GGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTCAAAGG 905  
Db 781 GGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTCAAAGG 840  
Qy 906 GAAGGTGAAGTGTGTGAGCTGAAGCCTGAGGTGGCGCCAGGTGACCTGTATCATCTCTGG 965  
Db 841 GAAGGTGAAGTGTGTGAGCTGAAGCCTGAGGTGGCGCCAGGTGACCTGTATCATCTCTGG 900  
Qy 966 CCAAGCGACCACTTTATGGGCAATGTGTCTCTGCTTCACTTCACTTGTGTGAAGCGGGA 1025  
Db 901 CCAAGCGACCACTTTATGGGCAATGTGTCTCTTCACTTCACTTGTGTGAAGCGGGA 960  
Qy 1026 GCGGGAACCTCAGGGAGGCGCTTCTTCTTCTGCGCATGGAACAGGCCCTTAAAGTGGC 1085  
Db 961 GCGGGAACCTCAGGGAGGCGCTTCTTCTTCTGCGCATGGAACAGGCCCTTAAAGTGGC 1020  
Qy 1086 GGAACGAGTTTCTGATTTGCGCGGAGCAACAGACCTCTGATCTTGAAGGACCAAGTCT 1145  
Db 1021 GGAACGAGTTTCTGATTTGCGCGGAGCAACAGACCTCTGATCTTGAAGGACCAAGTCT 1080  
Qy 1146 GAGCTGTGCTTCTCAGCCAGGCTTGGCAGCGAGAGGTGCTCCGGGATTTGCAAACTCTCT 1205  
Db 1081 GAGCTGTGCTTCTCAGCCAGGCTTGGCAGCGAGAGGTGCTCCGGGATTTGCAAACTCTCT 1140

Qy 1206 TCTCACCTGCCAAAGATGGAGAGAGTGCAGGAGACCCCTCAAGGAGGAGAGCGTCCAT 1265  
Db 1141 TCTCACCTGCCAAAGATGGAGAGAGTGCAGGAGACCCCTCAAGGAGGAGAGCGTCCAT 1200  
Qy 1266 ATCCAGGCGCATAGGACTTGCAGGTTTCTTAGAGCAGAGAGCATCTCCCATCGACGTCT 1325  
Db 1201 ATCCAGGCGCATAGGACTTGCAGGTTTCTTAGAGCAGAGAGCATCTCCCATCGACGTCT 1260  
Qy 1326 TTCTGCTCTTCTGGGAAATTTCTCACAACCTGGCAAGAGCATCTCCATCTTCTGTTCC 1385  
Db 1261 TTCTGCTCTTCTGGGAAATTTCTCACAACCTGGCAAGAGCATCTCCATCTTCTGTTCC 1320  
Qy 1386 ACTCTGCTCTGAGCAGCGCTGGGATGCTGAACTCTTTCAGAGAGATTTTATATAGAGAT 1445  
Db 1321 ACTCTGCTCTGAGCAGCGCTGGGATGCTGAACTCTTTCAGAGAGATTTTATATAGAGAT 1380  
Qy 1446 TTCTATAATTTTGTATACAAAGTCATGACTATCTCTAGAACTCTCTGTGGTTTGAATAATC 1505  
Db 1381 TTCTATAATTTTGTATACAAAGTCATGACTATCTCTAGAACTCTCTGTGGTTTGAATAATC 1440  
Qy 1506 ATTGAATTC 1514  
Db 1441 ATTGAATTC 1449

RESULT 10  
AR163452  
LOCUS AR163452 1300 bp DNA linear PAT 17-OCT-2001  
DEFINITION Sequence 6 from patent US 6270987.  
ACCESSION AR163452  
VERSION AR163452.1 GI:16234057  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 1300)  
AUTHORS Wang, Y. and Spellman, M.W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: US 6270987-A 6 07-AUG-2001;  
FEATURES Location/Qualifiers  
source  
1..1300  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 76.8%; Score 1163.4; DB 6; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 2.7e-284;  
Matches 1164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCCCGGGGCTCTCTGGGACCGCGGTTACCTGCTCTACTGCCCTGCATGGGGGC 60  
Db 136 ATGCCCGGGGCTCTCTGGGACCGCGGTTACCTGCTCTACTGCCCTGCATGGGGGC 195  
Qy 61 TTTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGGCTCTTGCAGCTTAAACCGT 120  
Db 196 TTTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGGCTCTTGCAGCTTAAACCGT 255  
Qy 121 ACCTTGGCTGCTCTCTGGGATGAGTACAGGATCAGAGCTCTTTCACCAACCTC 180  
Db 256 ACCTTGGCTGCTCTCTGGGATGAGTACAGGATCAGAGCTCTTTCACCAACCTC 315  
Qy 181 CATGTGCTCTTACCAAGATCACTTCAAGCTGAGGCCCTCCAGGGTTACCATCGGTCATC 240  
Db 316 CATGTGCTCTTACCAAGATCACTTCAAGCTGAGGCCCTCCAGGGTTACCATCGGTCATC 375  
Qy 241 AGCTTGGAGGATTTTCAATGGAGAGCTGGCAACCACTGGCCCCCTGAGAGCGGGTG 300  
Db 376 AGCTTGGAGGATTTTCAATGGAGAGCTGGCAACCACTGGCCCCCTGAGAGCGGGTG 435  
Qy 301 GCATACTCTTCTTGAAGTGGCAGCCAGCGAAGCCAGATAAGAGACGTGCCCCCATGAAG 360  
Db 436 GCATACTCTTCTTGAAGTGGCAGCCAGCGAAGCCAGATAAGAGACGTGCCCCCATGAAG 495







JOURNAL Patent: JP 2001527389-A 9 25-DEC-2001;  
GENENTECH INC  
COMMENT OS Unidentified  
PN JP 2001527389-A/9  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877  
PR 31-JAN-1997 US 08/792498,26-NOV-1997 US 08/978741 PI  
PC YANG WANG,MICHAEL W SPELLMAN  
C12N15/54,C12N9/10,C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
FH Key Location/Qualifiers  
FT source 1..1100  
FT /organism='Unidentified'.  
FEATURES Location/Qualifiers  
source 1..1100  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'

ORIGIN

Query Match 72.7%; Score 1100; DB 6; Length 1100;  
Best Local Similarity 100.0%; Pred.No. 3.4e-268;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGCGGGCTTCCTGGACCCGCGGTACTCTCTACTGCCCCCTGCATGGGGCG 60  
DB 1 ATGCCGCGGGCTTCCTGGACCCGCGGTACTCTCTACTGCCCCCTGCATGGGGCG 60

QY 61 TTGTGGAAACAGCGCCGATCACTTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120  
DB 61 TTGTGGAAACAGCGCCGATCACTTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

QY 121 ACCTTGCGTGTCCCTCTTGGATTGATGACCAGATCAACAAGCTCTTTTACCAACCTC 180  
DB 121 ACCTTGCGTGTCCCTCTTGGATTGATGACCAGATCAACAAGCTCTTTTACCAACCTC 180

QY 181 CATGTGTCTTACCAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTATC 240  
DB 181 CATGTGTCTTACCAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTATC 240

QY 241 AGCTTGAGGATTTTCATGGAAGCTGGCACCCACCCACTGCCCCTCTGAGAAGCGGTG 300  
DB 241 AGCTTGAGGATTTTCATGGAAGCTGGCACCCACCCACTGCCCCTCTGAGAAGCGGTG 300

QY 301 GCATACTGCTTTGAGTGGCGCCAGCCAGCCAGATTAAGAAGCTGCCCATGAAG 360  
DB 301 GCATACTGCTTTGAGTGGCGCCAGCCAGCCAGATTAAGAAGCTGCCCATGAAG 360

QY 361 GAAGAAAACCCCTTTGGCCCATCTTGGGATCAGTTTTTCAAGTTTCAACAAGTCGGAG 420  
DB 361 GAAGAAAACCCCTTTGGCCCATCTTGGGATCAGTTTTTCAAGTTTCAACAAGTCGGAG 420

QY 421 CTTTTFACAGGCAATTTCTTCAGTGCTTCTTACAGAGAACATGAGGACGAGATTTTCT 480  
DB 421 CTTTTFACAGGCAATTTCTTCAGTGCTTCTTACAGAGAACATGAGGACGAGATTTTCT 480

QY 481 CCNAAGAACATCCGCTGCTTGGCCCTCCAGAGCCCGCCAGCCAGTTCCTCGTCTAGAA 540  
DB 481 CCNAAGAACATCCGCTGCTTGGCCCTCCAGAGCCCGCCAGCCAGTTCCTCGTCTAGAA 540

QY 541 GAACACAGCCACTACAGAAGTACATGTPATGCTCAGACGAAATGGTCAAGACGGGAG 600  
DB 541 GAACACAGCCACTACAGAAGTACATGTPATGCTCAGACGAAATGGTCAAGACGGGAG 600

QY 601 GCCCAGATTATGCCCCACCTTGTCCGGCCCTATGTGGGCATTTCACTGCGCATTTGGCTCT 660  
DB 601 GCCCAGATTATGCCCCACCTTGTCCGGCCCTATGTGGGCATTTCACTGCGCATTTGGCTCT 660

QY 661 GACTGGAAGAACGCTGTGCCATGCTGAAGGACGGGATGACAGGCTGCGCATTTCAATGGCC 720  
DB 661 GACTGGAAGAACGCTGTGCCATGCTGAAGGACGGGATGACAGGCTGCGCATTTCAATGGCC 720



```

Db 430 GAAGGAACCCCTTTGGCCCATTTGGGATCAGTTTCATGTAGTTTCAACAAGTCGGAG 489
Qy 421 CTTTTCACAGGCAATTCCTTCAGTGCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 480
Db 490 CTTTTCACAGGCAATTCCTTCAGTGCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 549
Qy 481 CCAAGGAACATCCGGTGTGCTGCGAGAGCCCGAGGCTTCCCGTCTAGAA 540
Db 550 CCAAGGAACATCCGGTGTGCTGCGAGAGCCCGGCGCAGTCCCGTCTCGGAG 609
Qy 541 GAACACAGCCCACTACAGAGTACATGTTAGTGTGAGAGAAATGGTGAAGACGGAGAG 600
Db 610 GAACACAGCCCACTACAGAGTACATGTTAGTGTGAGAGAAATGGTGAAGACGGAGAG 669
Qy 601 GCCCAGATTCATGCCCATCTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTGGCTCT 660
Db 670 GCCCAGATTCATGCCCATCTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTGGCTCT 729
Qy 661 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACTGAGGCTCGCACTTCATGGCC 720
Db 730 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACCGCAGGCTCGCACTTCATGGCC 789
Qy 721 TCTCGGAGTGTGGGCTACAGCCGAGCAGCGGCCCCCTCACGATGACTATGTGC 780
Db 790 TCTCGGAGTGTGGGCTACAGCCGAGCAGCGGCCCCCTCACGATGACTATGTGC 849
Qy 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCC 840
Db 850 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCC 909
Qy 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db 910 CAGTCGGTCTATGTTGCTACTGATTCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 969
Qy 901 AAAGGGAAGGTGAAGGTGCTGAGCCTGAAGCCTGAGGTGGCCGAGCTCGATGATC 960
Db 970 AAAGGGAAGGTGAAGGTGCTGAGCCTGAAGCCTGAGGTGGCCGAGCTCGATGATC 1029
Qy 961 CTCGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTCACCTGCTTTGTGAAG 1020
Db 1030 CTCGCCAAGCTGACCACTTTATTGGCAACTGTGTCTCTCTCACCTGCTTTGTGAAG 1089
Qy 1021 CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTTTCGGCATGGAAGGCCCCCTTAAG 1080
Db 1090 CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTTTCGGCATGGAAGGCCCCCTTAAG 1149
Qy 1081 CTGCGGACGAGTTCTGA 1098
Db 1150 CTGCGGACGAGTTCTGA 1167

```

Search completed: November 9, 2005, 07:59:57  
 Job time : 6834.38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:23:13 ; Search time 883.842 Seconds  
(without alignments)  
10140.369 Million cell updates/sec

Title: US-09-774-954-1  
Perfect score: 1514  
Sequence: 1 ATGCCCGCGGCTCTCGGA.....TTTTGAAATCATTGAATTC 1514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04.\*  
1: geneseqn1980s.\*  
2: geneseqn1980s.\*  
3: geneseqn2000s.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002as.\*  
7: geneseqn2002bs.\*  
8: geneseqn2003as.\*  
9: geneseqn2003bs.\*  
10: geneseqn2003cs.\*  
11: geneseqn2003ds.\*  
12: geneseqn2004as.\*  
13: geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description         |
|------------|--------|-------------|--------|-------|---------------------|
| 1          | 1514   | 100.0       | 1514   | 2     | AAV65632 Human hea  |
| 2          | 1512.4 | 99.9        | 11284  | 2     | AAV65633 Plaemid c  |
| 3          | 1510.8 | 99.8        | 5218   | 13    | ACN40746 Tumour-as  |
| 4          | 1509.8 | 99.7        | 5266   | 10    | ADF81754 Leukaemia  |
| 5          | 1460.2 | 96.4        | 4230   | 4     | AAK51510 Human pol  |
| 6          | 1455.4 | 96.1        | 4850   | 4     | AAK52494 Human pol  |
| 7          | 1447.4 | 95.6        | 5009   | 2     | AAV65634 First Eco  |
| 8          | 854.8  | 56.5        | 915    | 4     | AAAD07629 Human sec |
| 9          | 608.6  | 40.2        | 10331  | 4     | Aak79915 Human imm  |
| 10         | 608.6  | 40.2        | 10331  | 4     | AAK03344 Human rep  |
| 11         | 442.8  | 29.2        | 477    | 9     | ACH35532 Human end  |
| 12         | 441    | 29.1        | 479    | 9     | ACH32225 Human end  |
| 13         | 406.2  | 26.8        | 610    | 13    | ADQ56824 Novel can  |
| 14         | 166.4  | 11.0        | 1209   | 4     | ABL13947 Drosophil  |
| c 15       | 150.4  | 9.9         | 3264   | 4     | ABL13946 Drosophil  |
| c 16       | 123.2  | 8.1         | 1545   | 4     | AAAF23894 Human sec |
| c 17       | 108.6  | 7.2         | 3793   | 4     | ABL13894 Drosophil  |
| c 18       | 71.8   | 4.7         | 3567   | 4     | ABL13952 Drosophil  |
| 19         | 60     | 4.0         | 60     | 6     | ABN44539 Human spl  |
| 20         | 56     | 3.7         | 760    | 3     | AAZ98210 Human sig  |

|      |      |     |        |    |           |           |
|------|------|-----|--------|----|-----------|-----------|
| 21   | 50   | 3.3 | 2000   | 8  | ADA71938  | Rice gene |
| c 22 | 46.2 | 3.1 | 2000   | 8  | ADA71938  | Rice gene |
| 23   | 42.4 | 2.8 | 1438   | 8  | ACA31561  | Prokaryot |
| c 24 | 41.2 | 2.7 | 114733 | 4  | AAAD08215 | Human gen |
| c 25 | 40.6 | 2.7 | 558    | 5  | AAH97782  | Murine 7- |
| 26   | 40.6 | 2.7 | 2036   | 13 | ADT44716  | Bacterial |
| 27   | 40.4 | 2.7 | 5019   | 13 | ADQ80232  | Brain-spe |
| 28   | 40.4 | 2.7 | 5704   | 12 | ADQ84111  | Human tum |
| 29   | 40.4 | 2.7 | 5932   | 10 | ADB75488  | Prostate  |
| 30   | 40   | 2.6 | 40     | 2  | AAV65637  | Primer am |
| 31   | 40   | 2.6 | 40     | 2  | AAV65635  | Northern  |
| 32   | 40   | 2.6 | 1554   | 8  | ADA69815  | Rice gene |
| 33   | 39   | 2.6 | 372    | 8  | ACA43606  | Prokaryot |
| 34   | 39   | 2.6 | 1079   | 10 | ADC37615  | Human nuc |
| 35   | 39   | 2.6 | 1131   | 11 | ABD01472  | Pseudomon |
| 36   | 39   | 2.6 | 1209   | 3  | AAA61286  | Human sec |
| 37   | 39   | 2.6 | 1209   | 8  | ADA40219  | Human sec |
| 38   | 39   | 2.6 | 1209   | 10 | ADD37708  | Human sec |
| 39   | 39   | 2.6 | 1209   | 10 | ADA56378  | Gene enco |
| 40   | 39   | 2.6 | 1353   | 11 | ABD01484  | Pseudomon |
| 41   | 39   | 2.6 | 1404   | 10 | ADC16737  | Human mod |
| 42   | 39   | 2.6 | 1528   | 4  | AAH15283  | Human CDN |
| 43   | 39   | 2.6 | 1555   | 4  | AAAS44695 | Human ful |
| 44   | 39   | 2.6 | 1566   | 4  | AAAS01567 | Human sec |
| 45   | 39   | 2.6 | 1575   | 6  | ADI16595  | Human NOV |

ALIGNMENTS

RESULT 1  
AAV65632  
ID AAV65632 standard; DNA; 1514 BP.  
XX  
AC AAV65632;  
XX  
DT 16-DEC-1998 (first entry)  
XX  
DE Human heart O-fucosyltransferase encoding DNA.  
XX  
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;  
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1..1100  
FT /tag= b  
FT /note= "this actively expressed O-fucosyltransferase  
FT sequence is claimed for in claim 9"  
FT CDS 1..1098  
FT /tag= a  
FT /product= "human heart O-fucosyltransferase"  
XX  
XX WO9833924-A1.  
XX  
XX  
XX PD 06-AUG-1998.  
XX  
XX PF 17-DEC-1997; 97WO-US023401.  
XX  
XX PR 31-JAN-1997; 97US-00792498.  
XX  
XX PR 26-NOV-1997; 97US-00578741.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX PI Wang Y, Spellman MW;  
XX  
XX DR WPI; 1998-437477/37.  
XX  
XX DR P-PSDB; AAW80571.  
XX  
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor  
XX domains - useful for diagnosis and treatment of diseases involving  
XX overexpression of the enzyme.







RESULT 3  
ACN40746  
ID ACN40746 standard; cDNA; 5218 BP.  
XX  
XX ACN40746;  
XX AC  
XX 18-NOV-2004 (first entry)  
XX  
XX Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.  
XX  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
XX Tumour; diagnosis; cell proliferative disorder; breast cancer;  
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
XX central nervous system cancer; bladder cancer; pancreatic cancer;  
XX cervical cancer; melanoma; leukaemia; hybridisation probe;  
XX chromosome identification; chromosome mapping; gene mapping;  
XX gene therapy; cytosstatic; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2004030615-A2.  
XX  
XX 15-APR-2004.  
XX  
XX 29-SEP-2003; 2003WO-US028547.  
XX  
XX 02-OCT-2002; 2002US-0414971P.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX Wu TD, Zhang Z, Zhou Y;  
XX  
XX WPI: 2004-347921/32.  
XX P-PSDB; AEM82223.  
XX  
XX New tumor-associated antigenic target polypeptides and nucleic acids,  
XX useful in preparing a medicament for treating or detecting a  
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
XX prostate cancer or tumor.  
XX  
XX Claim 1; SEQ ID NO 5725; 7273pp; English.  
XX  
XX The invention relates to human tumour-associated antigenic target (TAT)  
XX polypeptides, and their related nucleic acids. The TAT polypeptides are  
XX overexpressed in cancer tissues compared to normal tissues, and may thus  
XX serve as effective targets for the diagnosis and treatment of cancer in  
XX mammals. The invention also relates to nucleic acid and polypeptide  
XX sequences at least 80% identical to the TAT nucleic acids and  
XX polypeptides; expression vectors and host cells comprising a TAT nucleic  
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic  
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a  
XX TAT polypeptide; and methods and compositions for the treatment or  
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
XX antibodies, antagonists, binding molecules and compositions are useful  
XX for diagnosing or treating a cell proliferative disorder associated with  
XX increased TAT expression, particularly cancers such as breast cancer,  
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
XX cancer, pancreatic cancer, cervical cancer, cancers of the central  
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
XX used as hybridisation probes, in chromosome and gene mapping, in  
XX chromosome identification and in gene therapy. The present sequence  
XX represents a TAT nucleic acid of the invention  
XX  
XX Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;  
XX  
XX Query Match 99.8%; Score 1510.8; DB 13; Length 5218;  
XX Best Local Similarity 99.9%; Pred. No. 0;  
XX Matches 1512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX 1 ATGCCCGGGGCTCTGGGACCCGGGCTTACCTGCTTACTGCCCCGATGGGGCGC 60  
XX 119 ATGCCCGGGGCTCTGGGACCCGGGCTTACCTGCTTACTGCCCCGATGGGGCGC 178  
XX  
XX

QY 61 TTTGGAAACAGGCCGATCACTTCTTGGGCTCTCTGGCATTTTCAAGCTGTAAACCGT 120  
DB 179 TTTGGAAACAGGCCGATCACTTCTTGGGCTCTCTGGCATTTTCAAGCTGTAAACCGT 238  
QY 121 ACCTTGGCTGTCCCTCTCTTGGATTTGAGTACCAGCATCAAGGCTCTCTTTCACCAACCTC 180  
DB 239 ACCTTGGCTGTCCCTCTCTTGGATTTGAGTACCAGCATCAAGGCTCTCTTTCACCAACCTC 298  
QY 181 CATGTGCTCCTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCATC 240  
DB 299 CATGTGCTCCTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCATC 358  
QY 241 ACCTTGGAGGATTTTCATGGAGAGCTGSCACCCACCTGSCCTCAGAAAGCGGGTG 300  
DB 359 AGCTTGGAGGATTTTCATGGAGAGCTGSCACCCACCTGSCCTCAGAAAGCGGGTG 418  
QY 301 GCATCTGCTTGGAGTGGAGCCGAGGAGCCAGATGAAGAGCTGCCCATGAAG 360  
DB 419 GCATCTGCTTGGAGTGGAGCCGAGGAGCCAGATGAAGAGCTGCCCATGAAG 478  
QY 361 GAAGGAAACCCCTTTGGCCCATTCCTGGGATCAGTTTCAATGTAGTTTCAACAGTCGGAG 420  
DB 479 GAAGGAAACCCCTTTGGCCCATTCCTGGGATCAGTTTCAATGTAGTTTCAACAGTCGGAG 538  
QY 421 CTTTTCACAGGCAATTCCTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT 480  
DB 539 CTTTTCACAGGCAATTCCTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT 598  
QY 481 CCAAAGGAAACATCCGGTGTCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCGTCTAGAA 540  
DB 599 CCAAAGGAAACATCCGGTGTCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCGTCTAGAG 658  
QY 541 GAACACAGGCCACTACAGAGATGATGTTGTCAGACGAAATGGTGAAGACGGGAGAG 600  
DB 659 GAACACAGGCCACTACAGAGATGATGTTGTCAGACGAAATGGTGAAGACGGGAGAG 718  
QY 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGTCT 660  
DB 719 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGTCT 778  
QY 661 GACTGGAAGAACCGCTGTGCTGTCATGTGAAGACGGGACTGCGAGCTGCGACTTCATGGCC 720  
DB 779 GACTGGAAGAACCGCTGTGCTGTCATGTGAAGACGGGACTGCGAGCTGCGACTTCATGGCC 838  
QY 721 TCTCCGCAAGTGTGGGCTACAGCCGAGCAGCAGCGGCCCTCCAGATGATGTGTC 780  
DB 839 TCTCCGCAAGTGTGGGCTACAGCCGAGCAGCAGCGGCCCTCCAGATGATGTGTC 898  
QY 781 CTGCTGACCTGAAGAGGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGTGATGCC 840  
DB 899 CTGCTGACCTGAAGAGGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGTGATGCC 958  
QY 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGGTTATGTGCTGAGCTTCCAAACAGCTTTC 900  
DB 959 CAGTCGGTCTACGTTGCTACTGATTCGAGAGGTTATGTGCTGAGCTTCCAAACAGCTTTC 1018  
QY 901 AAAGGGAAGGTGAAGGTGGTGGCTGAGCCCTGAGGTGGCCAGGTCGAGCTGTACATC 960  
DB 1019 AAAGGGAAGGTGAAGGTGGTGGCTGAGCCCTGAAGCTGAGGTGGCCAGGTCGAGCTGTACATC 1078  
QY 961 CTCGGCCAGCGGACCACTTTATTTGGCAACTGTGCTCTCTCTTCTACTGCTTTGTGAAG 1020  
DB 1079 CTCGGCCAGCGGACCACTTTATTTGGCAACTGTGCTCTCTCTTCTACTGCTTTGTGAAG 1138  
QY 1021 CGGAGCGGAGCTTCAGGGGAGCGCTTCTTCTTCTCGCATGGACAGGCCCCCTTAG 1080  
DB 1139 CGGAGCGGAGCTTCAGGGGAGCGCTTCTTCTTCTCGCATGGACAGGCCCCCTTAG 1198  
QY 1081 CTGCGGAGCAGATTTCTGATTTCTGCGCGAGGACCAAGCCCTCTGATCTCTGGAGGACAG 1140  
DB 1199 CTGCGGAGCAGATTTCTGATTTCTGCGCGAGGACCAAGCCCTCTGATCTCTGGAGGACAG 1258

QY 1141 AGTCTGAGCTGCTCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGATGCAAACT 1200  
DB |||||  
1259 AGTCTGAGCTGCTCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGATGCAAACT 1318  
QY 1201 CTTCTTCTCACCTGCTCAAGATGGAGAGAGTGGCCAGGACCCCTCAAGAGGAGACGC 1260  
DB |||||  
1319 CTTCTTCTCACCTGCTCAAGATGGAGAGAGTGGCCAGGACCCCTCAAGAGGAGACGC 1378  
QY 1261 TCCATATCCAGGCGATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
DB |||||  
1379 TCCATATCCAGGCGATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1438  
QY 1321 GTGCTTCTGCTCTCTGCGAATTTCTACATGGCAAGCAGTCCAGCCTCCGCTTTCT 1380  
DB |||||  
1439 GTGCTTCTGCTCTCTGCGAATTTCTACACACTGGCAAGCAGTCCAGCCTCCGCTTTCT 1498  
QY 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACTCTTACAGAGATTTTTTATAGA 1440  
DB |||||  
1499 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACTCTTACAGAGATTTTTTATAGA 1558  
QY 1441 GAGATTTCTATTAATTTTATACAAAGTCAATGACTATCTTAGAATCTCTGTGTTTTTGA 1500  
DB |||||  
1559 GAGATTTCTATTAATTTTATACAAAGTCAATGACTATCTTAGAATCTCTGTGTTTTTGA 1618  
QY 1501 AAATCAATGAATTC 1514  
DB |||||  
1619 AAATCAATGAATTC 1632

## RESULT 4

ADF81754

ID ADF81754 standard; DNA; 5266 BP.

XX AC ADF81754;

XX DT 26-FEB-2004 (first entry)

XX DE Leukaemia-related DNA sequence #2310.

XX DX Cytostatic; Gene therapy; leukaemia; ss.

XX KW Unidentified.

XX OS

XX PN WO2003039443-A2.

XX PD 15-MAY-2003.

XX PF 04-NOV-2002; 2002WO-EP012303.

XX PR 05-NOV-2001; 2001EP-00126244.

XX PR 30-APR-2002; 2002EP-00009758.

XX PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX PA (UWLU-) UNIV LUDWIG MAXIMILIANS.

XX PA (HAFE/) HAFERLACH T.

XX PA (SCHO/) SCHOCH C.

XX PA (KERN/) KERN W.

XX PI Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;

XX PI Ellis R, Brors B, Mergenthaler S;

XX DR WPI; 2003-505037/47.

XX PT Determining the subtype of leukemia cells and whether a patient sample contains leukemia cells or other cells, useful for treating leukemia, comprises determining the expression profile of a group of markers in a patient sample.

XX PS Disclosure; SEQ ID NO 2310; 2938pp; English.

XX CC The present invention relates to a method (M1) for determining the subtype of leukaemia cells and whether a patient sample contains leukaemia cells. The method comprises determining the expression profile

CC of a group of markers in a patient sample. The method is useful for determining the presence of leukaemia cells, its types or subtypes, and CC for the preparation of a medicament for treating leukaemia.

XX

SQ Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;

Query Match 99.7%; Score 1509.8; DB 10; Length 5266;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1511; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCCCCGGGCTCTCGGAGCCCGCGGTACCTGCTCTACTGCCCCCTCGATGGGGCGC 60

DB 132 ATGCTGGGGCTCTCGGAGCCCGCGGTACCTGCTCTACTGCCCCCTCGATGGGGCGC 191

QY 61 TTGGGAACACAGGCGGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120

DB 192 TTGGGAACACAGGCGGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 251

QY 121 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTCAACCAACCTC 180

DB 252 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTCAACCAACCTC 311

QY 181 CATGTCTCTACAGAACTACTTCAAGCTGAGGCCCTCCAGGCTTACATCGGCTCATC 240

DB 312 CATGTCTCTACAGAACTACTTCAAGCTGAGGCCCTCCAGGCTTACATCGGCTCATC 371

QY 241 AGCTTGGAGGATTTTCATGGAGAGCTGGACCCACCCACTGGCCCCCTGAGAGCGGGTG 300

DB 372 AGCTTGGAGGATTTTCATGGAGAGCTGGACCCACCCACTGGCCCCCTGAGAGCGGGTG 431

QY 301 GCATACCTCTTGGAGTGGCCAGCCAGCAAGCCAGATTAAGAACAGCTGCCCATGAAG 360

DB 432 GCATACCTCTTGGAGTGGCCAGCCAGCAAGCCAGATTAAGAACAGCTGCCCATGAAG 491

QY 361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAAAGTCGGAG 420

DB 492 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAAAGTCGGAG 551

QY 421 CTTTTTACAGGCATTTCTTCTCAGTCTTCTTACAGAGAACATGAGGACGAGATTTTCT 480

DB 552 CTTTTTACAGGCATTTCTTCTCAGTCTTCTTACAGAGAACATGAGGACGAGATTTTCT 611

QY 481 CCAAAGAAACATCCGGTCTTCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAA 540

DB 612 CCAAAGAAACATCCGGTCTTCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAG 671

QY 541 GAAACACAGGCGCACTACAGAACTACATGATATGGTCAGACGAAATGGTGAAGACGGAGAG 600

DB 672 GAAACACAGGCGCACTACAGAACTACATGATATGGTCAGACGAAATGGTGAAGACGGAGAG 731

QY 601 GCCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTCATCTCGCATTTGGCTCT 660

DB 732 GCCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTCATCTCGCATTTGGCTCT 791

QY 661 GACTTGAAGAAGCGCTGTGCCATCTGAAGACCGGACTGCGAGGCTCCGCTTCTATGCCC 720

DB 792 GACTTGAAGAAGCGCTGTGCCATCTGAAGACCGGACTGCGAGGCTCCGCTTCTATGCCC 851

QY 721 TCTCCGAGTGTGTGGGTACAGCCGACAGCAGCGGCCCGCCCTCACGATGATATGTGC 780

DB 852 TCTCCGAGTGTGTGGGTACAGCCGACAGCAGCGGCCCGCCCTCACGATGATATGTGC 911

QY 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTCGTGTGATGCC 840

DB 912 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTCGTGTGATGCC 971

QY 841 CAGTCGGTCTACGTTTGTCTACTGATTCCGAGAGTTATGTGCCCTGAGCTCCACAGCTCTTC 900

DB 972 CAGTCGGTCTACGTTTGTCTACTGATTCCGAGAGTTATGTGCCCTGAGCTCCACAGCTCTTC 1031

QY 901 AAAGGAAGGTGAAGGTGGTGGAGCTGAAGCTCGAGGTGGCCCGAGCTCGACTGTACATC 960

DB 1032 AAAGGAAGGTGAAGGTGGTGGAGCTGAAGCTCGAGGTGGCCCGAGCTCGACTGTACATC 1091



Db 746 AGACGGGAGAGGCCAGATTCATGCCACCTTGTCCGCCCTATGTGGGCATTCATCTGC 805  
Qy 650 GCATTTGGCTCTGACTGGAAGAACGCTGTGCCATCTGAAGACGGGACTGCAGGCTCGC 709  
Db 806 GCATTTGGCTCTGACTGGAAGAACGCTGTGCCATCTGAAGACGGGACTGCAGGCTCGC 865  
Qy 710 ACTTCATGGCTCTCCGAGTGTGTGGGTACAGCCGACGACACGCGCCCTTCACGA 769  
Db 866 ACTTCATGGCTCTCCGAGTGTGTGGGTACAGCCGACGACACGCGCCCTTCACGA 925  
Qy 770 TGACTATGTGCTGCTGCTGACCTGAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGT 829  
Db 926 TGACTATGTGCTGCTGCTGACCTGAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGT 985  
Qy 830 CGCTGGATGCCAGTCCGCTCTACGTTGCTACTGATTTCCGAGAGTTATGTGCTCTGAGCTCC 889  
Db 986 CGCTGGATGCCAGTCCGCTCTACGTTGCTACTGATTTCCGAGAGTTATGTGCTCTGAGCTCC 1045  
Qy 890 AACAGCTCTTCAAGGGGAGGTGAAGTGTGTGAGCTTGAAGCCTGAGGTGGCCCAAGTCTG 949  
Db 1046 AACAGCTCTTCAAGGGGAGGTGAAGTGTGTGAGCTTGAAGCCTGAGGTGGCCCAAGTCTG 1105  
Qy 950 ACCTGTATCTCTCGGCCAAGCGACCACTTTATTTGGCAACTGTGTCTCTCTCTCACTG 1009  
Db 1106 ACCTGTATCTCTCGGCCAAGCGACCACTTTATTTGGCAACTGTGTCTCTCTCTCACTG 1165  
Qy 1010 CTTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGCTTCTTCTTCTTCCGCGATGGACA 1069  
Db 1166 CTTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGCTTCTTCTTCTTCCGCGATGGACA 1225  
Qy 1070 GGCCCCCTTAAGTGTGGGACGAGTTCTGATTTCTGSCCGGACACGACCCCTCTGATCCT 1129  
Db 1226 GGCCCCCTTAAGTGTGGGACGAGTTCTGATTTCTGSCCGGACACGACCCCTCTGATCCT 1285  
Qy 1130 GGAGGGACGAGTTGAGCTGTGCTTCTCCAGCCAGGCTGCGACCGAGAGTGTCTCGG 1189  
Db 1286 GGAGGGACGAGTTGAGCTGTGCTTCTCCAGCCAGGCTGCGACCGAGAGTGTCTCGG 1345  
Qy 1190 GATTGCAAACTCTCTTCTCACTCTGCCAAAGATGGAGAAGAGTGCCAGGACCCCTCAAG 1249  
Db 1346 GATTGCAAACTCTCTTCTCACTCTGCCAAAGATGGAGAAGAGTGCCAGGACCCCTCAAG 1405  
Qy 1250 GAGGAGAGCTTCATATCCAGGACATPAGGACTTGCAGGTTCTCTAGGAGCAGGACATC 1309  
Db 1406 GAGGAGAGCTTCATATCCAGGACATPAGGACTTGCAGGTTCTCTAGGAGCAGGACATC 1465  
Qy 1310 TCCCATCGACGCTGCTTCT 1369  
Db 1466 TCCCATCGACGCTGCTTCT 1525  
Qy 1370 CTCGCTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACCTCTTTCAGAGAT 1429  
Db 1526 CTCGCTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACCTCTTTCAGAGAT 1585  
Qy 1430 TTTTATPAGAGATTTCTATAATTTTGTATACAGGTCTGACTATCTCTAGAACTCTCT 1489  
Db 1586 TTTTATPAGAGATTTCTATAATTTTGTATACAGGTCTGACTATCTCTAGAACTCTCT 1645  
Qy 1490 GTGGTTTTTGAATCATGAATTC 1514  
Db 1646 GTGGTTTTTGAATCATGAATTC 1670

RESULT 6

AAK52494  
ID AAK52494 standard; cDNA; 4850 BP.  
XX AC  
XX AAK52494;  
XX AC  
XX 06-NOV-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 2023.  
XX

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
tissue growth factor; immunomodulatory; cancer; leukaemia;  
nervous system disorder; arthritis; inflammation; ss.  
Homo sapiens.  
WO200157190-A2.  
09-AUG-2001.  
05-FEB-2001; 2001WO-US004098.  
03-FEB-2000; 2000US-00496914.  
27-APR-2000; 2000US-00560875.  
20-JUN-2000; 2000US-00598075.  
19-JUL-2000; 2000US-00620325.  
01-SEP-2000; 2000US-00654936.  
15-SEP-2000; 2000US-00663561.  
20-OCT-2000; 2000US-00693325.  
30-NOV-2000; 2000US-00728422.  
(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
WPI; 2001-476283/51.  
P-ESDB; AAM79361.  
Nucleic acids encoding polypeptides with cytokine-like activities, useful  
in diagnosis and gene therapy.

Claim 1; Page 4424-4425; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the  
encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
cytokine cell proliferation or cell differentiation or which may induce  
production of other cytokines in other cell populations. The  
polynucleotides and polypeptides are useful in gene therapy, vaccines or  
peptide therapy. The polypeptides have various cytokine-like activities,  
e.g. stem cell growth factor activity, haematopoiesis regulating  
activity, tissue growth factor activity, immunomodulatory activity and  
activin/inhibin activity and may be useful in the diagnosis and/or  
treatment of cancer, leukaemia, nervous system disorders, arthritis and  
inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
(AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
sequence listing were missing at the time of publication

Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;

Query Match 96.1%; Score 1455.4; DB 4; Length 4850;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1459; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 50 GCATGGGGCGCTTGGGAACAGCCGATCATTCTTTGGGCTCTCTGGCATTTTCCAAAGC 109  
Db 206 GCCAGGGGCGCTTGGGAACAGCCGATCATTCTTTGGGCTCTCTGGCATTTTCCAAAGC 265  
Qy 110 TGCTAAACCGGTACCTTGGCTGTCTCTCTTTGGATTGAGTACCAGCATCACAAAGCTCTCTT 169  
Db 266 TGCTGAACCGTCTCTTGGCTGTCTCTCTTTGGATTGAGTACCAGCATCACAAAGCTCTCTT 325  
Qy 170 TCACCAACCTCCATGTCTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACC 229  
Db 326 TCACCAACCTCCATGTCTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACC 385  
Qy 230 ATCGGTCATCAGCTTGGAGGATTTTCATGAGAGCTGGCACCACCCACCTGGCCCCCTG 289  
Db 386 ATCGGTCATCAGCTTGGAGGATTTTCATGAGAGCTGGCACCACCCACCTGGCCCCCTG 445  
Qy 290 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCAGCAAGCCAGATGAAGAAGCGT 349

Db 446 AGAAGCGGTTGGCATATCTGTTTGGGTGGAGCCAGCCAGATAGAGAGCGT 505  
Qy 350 GCGCCATGAAAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCA 409  
Db 506 GCGCCATGAAAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCA 565  
Qy 410 ACAAGTGGAGCTTTTACAGCAATTTCTTCACTGCTTCTTACAGAGAAACAATGGAGCC 469  
Db 566 ACAAGTGGAGCTTTTACAGCAATTTCTTCACTGCTTCTTACAGAGAAACAATGGAGCC 625  
Qy 470 AGAGATTTTCTCCAAAGAGACATCGGTGCTTCCCTGCGCAGAGCCCGCCAGCTTCC 529  
Db 626 AGAGATTTTCTCCAAAGAGACATCGGTGCTTCCCTGCGCAGAGCCCGCCAGCTTCC 685  
Qy 530 CCGCTAGAGAAACACAGGCCCATACAGAGATCATGTGATGTGTCAGACGAATGTGTA 589  
Db 686 CCGCTAGAGAAACACAGGCCCATACAGAGATCATGTGATGTGTCAGACGAATGTGTA 745  
Qy 590 AGACGGGAGAGCCACAGATTATGATGCCCATCTGTGCTCCGGCCCTATGTGGCAATCATCTGC 649  
Db 746 AGACGGGAGAGCCACAGATTATGATGCCCATCTGTGCTCCGGCCCTATGTGGCAATCATCTGC 805  
Qy 650 GCATTGGCTCTGACTGGAAGAACCGCTGTGCCATGCTGAAAGACGGGACTGCAGGCTCGC 709  
Db 806 GCATTGGCTCTGACTGGAAGAACCGCTGTGCCATGCTGAAAGACGGGACTGCAGGCTCGC 865  
Qy 710 ACTTCATGGCTCTCCGCACTGTGTGGCTACAGCCGAGCAGAGCGGCCCCCTCAGCA 769  
Db 866 ACTTCATGGCTCTCCGCACTGTGTGGCTACAGCCGAGCAGAGCGGCCCCCTCAGCA 925  
Qy 770 TGCACTATGCTGCTCCCTGACCTGAAAGAGATCCAGAGGCTGTGAAGCTCTCGGTGAGGT 829  
Db 926 TGCACTATGCTGCTCCCTGACCTGAAAGAGATCCAGAGGCTGTGAAGCTCTCGGTGAGGT 985  
Qy 830 CGCTGGATCCCGAGTCCGTCTAGTTGCTACTGATTCGAGAGTTATGCTGAGCTCC 889  
Db 986 CGCTGGATCCCGAGTCCGTCTAGTTGCTACTGATTCGAGAGTTATGCTGAGCTCC 1045  
Qy 890 AACAGCTTTCAAAGGAAAGGTGAAGTGTGAGCTGTGAAGCTGAGGTGGCCCAAGTTCG 949  
Db 1046 AACAGCTTTCAAAGGAAAGGTGAAGTGTGAGCTGTGAAGCTGAGGTGGCCCAAGTTCG 1105  
Qy 950 ACTGTATCATCTCGGCCAAGCCGACACTTTATTTGGCACTGTGTCTCTCTTCACTG 1009  
Db 1106 ACTGTATCATCTCGGCCAAGCCGACACTTTATTTGGCACTGTGTCTCTCTTCACTG 1165  
Qy 1010 CTTTGTGAAGCGGAGCGGAGCTCCAGGGAGGCGCTCTTCTTCTTGGCATGGACA 1069  
Db 1166 CTTTGTGAAGCGGAGCGGAGCTCCAGGGAGGCGCTCTTCTTCTTGGCATGGACA 1225  
Qy 1070 GCGCCCTAAGCTCGGACAGTTCGTATTTCTGGCCGAGCACAGACCTCTGATCCT 1129  
Db 1226 GCGCCCTAAGCTCGGACAGTTCGTATTTCTGGCCGAGCACAGACCTCTGATCCT 1285  
Qy 1130 GAGGAGACAGAGTCTGAGTGTGCTTCCAGCCAGGCTCGCAGCCAGAGTGTCTCCGG 1189  
Db 1286 GAGGAGACAGAGTCTGAGTGTGCTTCCAGCCAGGCTCGCAGCCAGAGTGTCTCCGG 1345  
Qy 1190 GATTGAAACTCTCTTCTACCTGCGCAAGATGAGAGATGCCAGGAGCCCTCAG 1249  
Db 1346 GATTGAAACTCTCTTCTACCTGCGCAAGATGAGAGATGCCAGGAGCCCTCAG 1405  
Qy 1250 GAGGAGACAGCTCCATATCCAGGGCATAGGACTTGCAGGTTCTTAGAGCAGAGCATC 1309  
Db 1406 GAGGAGACAGCTCCATATCCAGGGCATAGGACTTGCAGGTTCTTAGAGCAGAGCATC 1465  
Qy 1310 TCCCATCGCAGCTGCTTTCTGCTTCTGCGAAATTTCTCACACTGGCAAGAGCTCCAGC 1369  
Db 1466 TCCCATCGCAGCTGCTTTCTGCTTCTGCGAAATTTCTCACACTGGCAAGAGCTCCAGC 1525  
Qy 1370 CTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAATCTTTCAGAGAT 1429

Db 1526 CTCGCTCTTCTGGTCCACTCTGCTCTGAGAGCCTGGATGCTGAACCTTTCAGAGAGAT 1585  
Qy 1430 TTTTATATAGAGAGATTTCTATAATTTTATACAGGTCAAGTCACTATCTAGAACTCTCT 1489  
Db 1586 TTTTATATAGAGAGATTTCTATAATTTTATACAGGTCAAGTCACTATCTAGAACTCTCT 1645  
Qy 1490 GTGGTTTTTGAAATCATTTGAATTC 1514  
Db 1646 GTGGTTTTTGAAATCATTTGAATTC 1670

RESULT 7  
AAV5634  
ID AAV5634 standard; cDNA; 5009 BP.  
XX  
AC AAV5634;  
XX  
XX 16-DEC-1998 (first entry)  
XX  
DE First EcoRI nucleotide fragment of human KIAA0180.  
XX  
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;  
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO9833924-A1.  
XX  
PD 06-AUG-1998.  
XX  
PF 17-DEC-1997; 97WO-US023401.  
XX  
PR 31-JAN-1997; 97US-00792498.  
PR 26-NOV-1997; 97US-00978741.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX Wang Y, Spellman MW;  
XX  
XX WPI; 1998-437477/37.  
DR P-PSDB; AAN80577.  
XX  
PT Human O-fucosyltransferase able to glycosylate epidermal growth factor  
PT domains - useful for diagnosis and treatment of diseases involving  
PT overexpression of the enzyme.  
XX  
PS Example 1; Fig 11; 90pp; English.  
XX  
CC This represents a first EcoRI nucleotide fragment of human KIAA0180. This  
CC 5009 basepairs partial cDNA encodes for a protein of unknown function  
CC from myelast cel line KG-1. The invention provides a human heart O-  
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)  
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of  
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF  
CC domains, are used in diagnosis and treatment of conditions associated  
CC with overexpression of O-fucosyltransferase, to promote survival of  
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are  
CC used to detect gene amplification and expression. The expression can also  
CC be determined at the protein level using antibodies specific for O-  
CC fucosyltransferase  
XX  
SQ Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Query Match 95.6%; Score 1447.4; DB 2; Length 5009;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 66 GAACAGGCGGATCACTTCTTGGGCTCTGGAGTTTGCAGAGCTGTAAACCGTACCTT 125  
Db 1 GAACAGGCGGATCACTTCTTGGGCTCTGGAGTTTGCAGAGCTGTAAACCGTACCTT 60  
Qy 126 GGCTGTCCCTCTCGATTTGATGATACAGATCAAGCCCTCTTTCACCACTCCATGT 185







14-JUL-2000; 2000US-0218290P.  
26-JUL-2000; 2000US-0220963P.  
26-JUL-2000; 2000US-0220964P.  
14-AUG-2000; 2000US-0224518P.  
14-AUG-2000; 2000US-0224519P.  
14-AUG-2000; 2000US-0225213P.  
14-AUG-2000; 2000US-0225214P.  
14-AUG-2000; 2000US-0225266P.  
14-AUG-2000; 2000US-0225267P.  
14-AUG-2000; 2000US-0225268P.  
14-AUG-2000; 2000US-0225270P.  
14-AUG-2000; 2000US-0225447P.  
14-AUG-2000; 2000US-0225757P.  
14-AUG-2000; 2000US-0225758P.  
14-AUG-2000; 2000US-0225759P.  
18-AUG-2000; 2000US-0226279P.  
22-AUG-2000; 2000US-0226681P.  
22-AUG-2000; 2000US-0226688P.  
22-AUG-2000; 2000US-0227182P.  
23-AUG-2000; 2000US-0227009P.  
30-AUG-2000; 2000US-0228924P.  
01-SEP-2000; 2000US-0229287P.  
01-SEP-2000; 2000US-0229343P.  
01-SEP-2000; 2000US-0229344P.  
01-SEP-2000; 2000US-0229345P.  
05-SEP-2000; 2000US-0229509P.  
05-SEP-2000; 2000US-0229513P.  
06-SEP-2000; 2000US-0230437P.  
06-SEP-2000; 2000US-0230438P.  
08-SEP-2000; 2000US-0231242P.  
08-SEP-2000; 2000US-0231243P.  
08-SEP-2000; 2000US-0231244P.  
08-SEP-2000; 2000US-0231413P.  
08-SEP-2000; 2000US-0231414P.  
08-SEP-2000; 2000US-0232080P.  
08-SEP-2000; 2000US-0232081P.  
12-SEP-2000; 2000US-0231968P.  
14-SEP-2000; 2000US-0232397P.  
14-SEP-2000; 2000US-0232398P.  
14-SEP-2000; 2000US-0232399P.  
14-SEP-2000; 2000US-0232400P.  
14-SEP-2000; 2000US-0232401P.  
14-SEP-2000; 2000US-0233063P.  
14-SEP-2000; 2000US-0233064P.  
14-SEP-2000; 2000US-0233065P.  
21-SEP-2000; 2000US-0234223P.  
21-SEP-2000; 2000US-0234274P.  
25-SEP-2000; 2000US-0234997P.  
25-SEP-2000; 2000US-0234998P.  
26-SEP-2000; 2000US-0235484P.  
27-SEP-2000; 2000US-0235834P.  
27-SEP-2000; 2000US-0235836P.  
29-SEP-2000; 2000US-0236327P.  
29-SEP-2000; 2000US-0236327P.  
29-SEP-2000; 2000US-0236367P.  
29-SEP-2000; 2000US-0236368P.  
29-SEP-2000; 2000US-0236369P.  
29-SEP-2000; 2000US-0236370P.  
02-OCT-2000; 2000US-0236802P.  
02-OCT-2000; 2000US-0237037P.  
02-OCT-2000; 2000US-0237038P.  
02-OCT-2000; 2000US-0237039P.  
02-OCT-2000; 2000US-0237040P.  
13-OCT-2000; 2000US-0239935P.  
13-OCT-2000; 2000US-0239937P.  
20-OCT-2000; 2000US-0240960P.  
20-OCT-2000; 2000US-0241221P.  
20-OCT-2000; 2000US-0241785P.  
20-OCT-2000; 2000US-0241786P.  
20-OCT-2000; 2000US-0241787P.  
20-OCT-2000; 2000US-0241808P.  
20-OCT-2000; 2000US-0241809P.  
20-OCT-2000; 2000US-0241826P.  
01-NOV-2000; 2000US-0244617P.  
08-NOV-2000; 2000US-0246474P.  
08-NOV-2000; 2000US-0246475P.  
08-NOV-2000; 2000US-0246476P.  
08-NOV-2000; 2000US-0246477P.  
08-NOV-2000; 2000US-0246478P.  
08-NOV-2000; 2000US-0246523P.  
08-NOV-2000; 2000US-0246524P.  
08-NOV-2000; 2000US-0246525P.  
08-NOV-2000; 2000US-0246526P.  
08-NOV-2000; 2000US-0246527P.  
08-NOV-2000; 2000US-0246528P.  
08-NOV-2000; 2000US-0246532P.  
08-NOV-2000; 2000US-0246609P.  
08-NOV-2000; 2000US-0246610P.  
08-NOV-2000; 2000US-0246611P.  
08-NOV-2000; 2000US-0246613P.  
17-NOV-2000; 2000US-0249207P.  
17-NOV-2000; 2000US-0249208P.  
17-NOV-2000; 2000US-0249209P.  
17-NOV-2000; 2000US-0249210P.  
17-NOV-2000; 2000US-0249211P.  
17-NOV-2000; 2000US-0249212P.  
17-NOV-2000; 2000US-0249213P.  
17-NOV-2000; 2000US-0249214P.  
17-NOV-2000; 2000US-0249215P.  
17-NOV-2000; 2000US-0249216P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249218P.  
17-NOV-2000; 2000US-0249244P.  
17-NOV-2000; 2000US-0249245P.  
17-NOV-2000; 2000US-0249264P.  
17-NOV-2000; 2000US-0249265P.  
17-NOV-2000; 2000US-0249297P.  
17-NOV-2000; 2000US-0249299P.  
17-NOV-2000; 2000US-0249300P.  
01-DEC-2000; 2000US-0250160P.  
01-DEC-2000; 2000US-0250391P.  
05-DEC-2000; 2000US-0251030P.  
05-DEC-2000; 2000US-0251988P.  
05-DEC-2000; 2000US-0256719P.  
06-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.  
08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.  
05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-483426/52.  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and metastasis.  
Disclosure; SEQ ID NO 34727; 3071pp + Sequence Listing; English.  
AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For  
example, they may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of (I) by expressing inactive proteins or to  
supplement the patients own production of (I). Additionally, (I)  
polynucleotides may be used to produce the secreted (I), by inserting the  
nucleic acids into a host cell and culturing the cell to express the  
protein. (I) proteins and polynucleotides may be used to prevent,  
diagnose and treat immune/hematopoietic-related diseases, especially



PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241828P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.

XX PA  
XX PI Rosen CA, Baraesh SC, Ruben SM;  
XX WPI; 2001-465570/50.  
DR

XX Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.  
PT  
XX  
PS Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.  
XX  
CC The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, CC including cancer. The present sequence is a genomic sequence encoding a CC protein of the invention  
XX  
SQ Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;  
Query Match 40.2%; Score 608.6; DB 4; Length 10331;  
Best Local Similarity 98.6%; Pred. No. 4e-166;  
Matches 614; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 892 CAGCTCTTCAAGGGAAGGTGAAGTGTGAGCTGAGCTGAAGCTGAGGTGGCCAGGTCCAC 951  
DB 6127 CTGTGCTTCTTCTCTGAGGTGAAGTGTGAGCTGAGCTGAGGTGGCCAGGTCCAC 6186  
QY 952 CTGTACATCTCTGCGCCAAAGCCGACACTTTATTTGGCAACTGTCTCTCTTCACTGCC 1011  
DB 6187 CTGTACATCTCTGCGCCAAAGCCGACACTTTATTTGGCAACTGTCTCTCTTCACTGCC 6246  
QY 1012 TTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGGTCTTCTTTTTCGGCATGGACAGG 1071  
DB 6247 TTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGGTCTTCTTTTTCGGCATGGACAGG 6306  
QY 1072 CCCCCTAAGCTGCGGGACGAGTTCGTATTTCTGGCCGAGCACACAGCCTCTGATCCTGG 1131  
DB 6307 CCCCCTAAGCTGCGGGACGAGTTCGTATTTCTGGCCGAGCACACAGCCTCTGATCCTGG 6366  
QY 1132 AGGACCCAGAGTCTGAGCTGGTCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGA 1191  
DB 6367 AGGACCCAGAGTCTGAGCTGGTCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGA 6426  
QY 1192 TTGCAAACTCTCTTCTCAGCTGCAAGATGGAGAGTGGCCAGGACCCCTCAAGGA 1251  
DB 6427 TTGCAAACTCTCTTCTCAGCTGCAAGATGGAGAGTGGCCAGGACCCCTCAAGGA 6486  
QY 1252 GGGAGACGCTCCATATCCAGGGCATAGGACTTTCAGGTTCCTAGGAGCAGGACATCTC 1311  
DB 6487 GGGAGACGCTCCATATCCAGGGCATAGGACTTTCAGGTTCCTAGGAGCAGGACATCTC 6546  
QY 1312 CCATCGCACGCTGCTTCTGCTTCTGGGAATTTCTCACAATGGCAAGCAGTCCAGCCT 1371  
DB 6547 CCATCGCACGCTGCTTCTGCTTCTGGGAATTTCTCACAATGGCAAGCAGTCCAGCCT 6606  
QY 1372 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTCAGAGAGATTT 1431  
DB 6607 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTCAGAGAGATTT 6666  
QY 1432 TTTTATAGAGAGATTTCTATAATTTTATCAAGGTTCATGACTATCTTAGAATCTCTCTGT 1491  
DB 6667 TTTTATAGAGAGATTTCTATAATTTTATCAAGGTTCATGACTATCTTAGAATCTCTCTGT 6726  
QY 1492 GGTTTTGAATAATCATTGAATTC 1514  
DB 6727 GGTTTTGAATAATCATTGAATTC 6749  
RESULT 11  
ACH35532  
ID ACH35532 standard; cDNA; 477 BP.  
XX  
XX AC ACH35532;  
XX  
XX DT 13-OCT-2003 (first entry)  
XX DE Human endothelial cell cDNA #3665.  
XX



| Query Match           | 29.1%;  | Score 441;          | DB 9;     | Length 479; |
|-----------------------|---|---------------------|-----------|-------------|
| Best Local Similarity | 98.9%;  | Pred. No. 6.9e-118; |           |             |
| Matches 444;          | Conservative 0;   | Mismatches 5;       | Indels 0; | Gaps 0;     |
| 868                   | GAGAGTTATGTGCGCTGAGCTCCAAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCGCTG | 927                 |           |             |
| 31                    | GAGAGTTATGTGCGCTGAGCTCCACACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCGCTG | 90                  |           |             |
| 928                   | AAGCCTGAGGTGGCCCGCCAGGTGCACTGTACATCTCTGGCCCAAGCCGACCACTTTATTGGC | 987                 |           |             |
| 91                    | AAGCCTGAGGTGGCCCGCCAGGTGCACTGTACATCTCTGGCCCAAGCCGACCACTTTATTGGC | 150                 |           |             |
| 988                   | AACGTGTGTCCTCTCTTTCACCTTTGTGTAAGCGGAGCGGAGACCTCCAGGGGAGGCGCG    | 1047                |           |             |
| 151                   | AACGTGTGTCCTCTCTTTCACCTTTGTGTAAGCGGAGCGGAGACCTCCAGGGGAGGCGCG    | 210                 |           |             |
| 1048                  | TCCTCTTTCTTCGGCATGGACAGGCCCCCTTAAGCTGCGGGACGAGTTCTGATTCCTGGCGG  | 1107                |           |             |
| 211                   | TCCTCTTTCTTCGGCATGGACAGGCCCCCTTAAGCTGCGGGACGAGTTCTGATTCCTGGCGG  | 270                 |           |             |
| 1108                  | GAGCACAGACCTCTGTATCTGTGAGGGAACAGAGTCTGAGCTGGTCTTCTCCAGCCAGGC    | 1167                |           |             |
| 271                   | GAGCACAGACCTCTGTATCTGTGAGGGAACAGAGTCTGAGCTGGTCTTCTCCAGCCAGGC    | 330                 |           |             |
| 1168                  | CTGGCAGCCAGAGTGTCTCCGGGATTGCAAACTCCTCTTCTCACCTTCGCAAGATGAGA     | 1227                |           |             |
| 331                   | CTGGCAGCCAGAGTGTCTCCGGGATTGCAAACTCCTCTTCTCACCTTCGCAAGATGAGA     | 390                 |           |             |
| 1228                  | AGAGTGCCAGGGAACCCCTCAAGGAGGAGAGCGCTCCATATCCCAAGGCATAGGACTTGCA   | 1287                |           |             |
| 391                   | AGAGTGCCAGGGAACCCCTCAAGGAGGAGAGCGCTCCATATCCCAAGGCATAGGACTTGGA   | 450                 |           |             |
| 1288                  | GGTTCTTAGGAGCAGGAGCATCTCCCATC                                   | 1316                |           |             |
| 451                   | GGTTCTTAGGAGCAGGAGCATCTCCCATC                                   | 479                 |           |             |

SULT 13  
Q56824  
ADQ56824 standard; DNA; 610 BP.  
ADQ56824;  
21-OCT-2004 (first entry)  
Novel canine microarray-related DNA sequence SeqID8126.  
canine microarray; drug screening; toxicity assay;  
environmental pollutant; cellular response; gene expression profile;  
toxic response; liver necrosis; fatty liver disease;  
protein adduct formation; hepatitis; dog; ds.  
Canis familiaris.  
WO2004063324-A2.  
29-JUL-2004.  
05-MAY-2003; 2003WO-US013853.  
03-MAY-2002; 2002US-0377240P.  
(GENE-) GENE LOGIC INC.  
(PFIZ ) PFIZER PROD INC.  
Diggins JC, Porter M, Wei T;  
WPI; 2004-561890/54.  
New isolated nucleic acid molecule, useful for drug screening and  
toxicity assays or for assessing the impact, including toxicity, of a  
compound, pharmaceutical agent or environmental pollutant on a cell or

living organism.

Claim 1; SEQ ID NO 8126; 41pp; English.

This invention is related to a novel isolated canine nucleic acid sequences and the construction of canine microarrays containing a significant portion of the canine genome. The isolated canine nucleic acid sequences of the invention may be useful for drug screening and toxicity assays. The invention is therefore useful for assessing the impact, including toxicity, of a compound, pharmaceutical agent or environmental pollutant on a cell or living organism. The methods are useful for detecting genes that are up- or down-regulated in canines in a disease state. The sequences are useful as diagnostic agents or markers to detect a cellular response in a sample individually or as part of a gene expression profile. It is also useful as a target for agents that modulate gene expression or activity. The database is useful for producing electronic Northernblots that allow the user to determine the cell type or tissue in which a given gene is expressed and to allow determination of the abundance or expression level of a given gene in a particular tissue or cell. The methods are useful for determining the similarity of a toxic response to one or more individual compounds. The methods are useful for predicting at least one toxic response or the likelihood that a compound or test agent will induce various specific pathologies such as those of the liver (liver necrosis, fatty liver disease, protein adduct formation or hepatitis), those of the kidney, heart, brain or testes, or other pathologies associated with at least one of the toxins. The methods are also useful for predicting or elucidating the potential cellular pathways influenced, induced or modulated by the compound or test agent due to the similarity of the expression profile is compared to the profile induced by a known toxin. The present sequence is that of a canine DNA sequence which was claimed for use during the production of a canine microarray of the invention.

| Seq | Sequence              | 610 BP  | 155 A        | 171 C    | 149 G      | 129 T  | 0 U    | 6 Other |
|-----|-----------------------|---|--------------|----------|------------|--------|--------|---------|
|     | Query Match           | 26.8%   | Score        | 406.2    | DB 13      | Length | 610    |         |
|     | Best Local Similarity | 89.0%   | Pred. No.    | 1.1e-107 |            |        |        |         |
|     | Matches               | 446   | Conservative | 0        | Mismatches | 54     | Indels | 1       |
|     |                       |   |              |          |            |        |        | Gaps    |
| Qy  | 171                   | CACCAACCTCCATGTGTCCTACCAAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCA    | 230          |          |            |        |        |         |
| Db  | 1                     | CACCAATCTCCATGTGTGCTTACCAGAAGTACTTCAAGCTNNAGCCCTCAGGNNNTACCA    | 60           |          |            |        |        |         |
| Qy  | 231                   | TCGGGTTCATCAGCTTGGAGGATTTCTATGGAGAGCTGGCACCCACCCACTGGGCCCTCGA   | 290          |          |            |        |        |         |
| Db  | 61                    | TCGGGTTCATCAGCTTATAGAGACTTCTATGGAGAAGCTGGCACCCACTCCTGGGCCCTCCGA | 120          |          |            |        |        |         |
| Qy  | 291                   | GAAGCGGGTGGCATACTGCTTTCAGGTGGCAGCCACAGCGAAGCCACAGATAAGAAGACGTG  | 350          |          |            |        |        |         |
| Db  | 121                   | GAGCGAGTGGCATACTGCTTTGAAGTGGCAGCCACAGCGAAGCCCTGATAGAAGACATG     | 180          |          |            |        |        |         |
| Qy  | 351                   | CCGCATGAAGGAAGGAAACCCCTTTGGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAA  | 410          |          |            |        |        |         |
| Db  | 181                   | CCCATGAAGGAAGGAAATCCCTTTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAA     | 240          |          |            |        |        |         |
| Qy  | 411                   | CAAGTCGGAGCTTTTTTACAGGCATTTCTTTCAAGTGTCTCTACAGAGAAACAATGGAGCCA  | 470          |          |            |        |        |         |
| Db  | 241                   | CAAGTCCAGCTTTTTTTCAGGCATTTCTTTCAAGTGTCTCTACAGAGAAACAATGGATTC    | 300          |          |            |        |        |         |
| Qy  | 471                   | GAGATTTTCTCAAAGGAACATCCGGTGTCTTGGCCCTGCCACAGAGACCCACGCCAGTTCCC  | 530          |          |            |        |        |         |
| Db  | 301                   | GAGATTTTCTCAAAGGAACATCCAGTGTCTGGGCC-CCCGGGAGCCGCCCCAGTTCCC      | 359          |          |            |        |        |         |
| Qy  | 531                   | CGTCTTAGAGAAACACAGGCCACTACAGAGTACATGGTATGTCACAGCAAAATGGTAA      | 590          |          |            |        |        |         |
| Db  | 360                   | CGTCTTAGAGAGACAGACCACTCAGAGATGATGTTGATGTCAGATGAGATGGTGAG        | 419          |          |            |        |        |         |
| Qy  | 591                   | GACGGAGAGGCCACAGATTATGCCCACTTGTCCGGCCCTATGTGGGGCATTCATCTCGCG    | 650          |          |            |        |        |         |
| Db  | 420                   | GACGGAGAGGCCCAATTCATCCACCACTCATCCGGCCGTACGTGGGCATTCACCTGCG      | 479          |          |            |        |        |         |
| Qy  | 651                   | CATTGGCTCTGACTGGGAAGAA  | 671          |          |            |        |        |         |



XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB69843.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

XX Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics, and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA  
CC sequences (ABLI01840-ABLI16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Query Match 9.9%; Score 150.4; DB 4; Length 3264;  
Best Local Similarity 50.9%; Pred. No. 1.4e-32;  
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;

QY 49 TGCATGGGGCGCTTTGGGAACCCAGCCGATCATCTTTGGGCTCTCTGGCATTTGSCAAAG 108  
DB 2098 TTCAAAGGAGCGCTTTGGCAACCCAGCCGACCATCTTCTGGGATCATTTGGCCCTTCGCCAAG 2039  
QY 109 CTGCTAAACCGTACTTGGCTGTCTCCCTCTGGATTGATGATCCAGCATCAAGCTCCT 168  
DB 2038 GCGCTTAATCGCACCCCTGATCTGCGCCGCTGGGTGGAGT-----ATCGTAGGGGTGAA 1985  
QY 169 TTCACCAACCTCCATGTGTCTTACCAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTAC 228  
DB 1984 CTGGATCCCGTCAGGTACCGTTCAACACATATTTGAAGTGGAGCCCTCGAAGGAATAC 1925  
QY 229 CATCGGGTCATCAGCTTGGAGGATTTTCATGAGAGAGCTGGCACCCACCCACTTGGCCCT 288  
DB 1924 CATCGCGTCATCACCATGCGAGATTTTCATGCGACCTTGGCGAGCATTTGGCCAGAA 1865  
QY 289 GAGAGCGGGTGGCATATCTGCTTTGAGG-----TGGCAGCCCGAGCGAAGC 333  
DB 1864 TCGGAGCGAGTGTCTATTTTGTCTACAAGGAACGATATAGCTTTCAGCAGGAGAAGAACGAT 1805  
QY 334 CCAGATAGAGAGCTGTGCGCATGAGGAGGAACCCCTTTGGCCATCTGGGATCAG 393  
DB 1804 CCAGACAGCCCAATTTGCCACGCCAAGGATGGCAATCTTTTGGTCCCTTTTGGACACT 1745  
QY 394 TTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTTCAGGCAATTTCTCAGTGTCTCTAC 453  
DB 1744 TTTTCATAGACTTTTGTGGGTGAGGTTCTATGCGCCACTTCTATTTGATGTGCATCAT 1685  
QY 454 AG-----AGAACATGAGGCGCAGAGATTTTCTCCAAAGGAACATCCGGTGTCTGCC 504  
DB 1684 AGCAACGAGGCTGCCAAGTGGCAGACCAATATCTCGAGAATCATATCCCGTACTCGCG 1625

QY 505 CTGCCAGGAGCCCGAGCCAGTTCCCGTCTCTAGAGAACACAGSCCACTACAGAAGTAC 564  
DB 1624 TTCACCGAGGCTCCGGCTAGTTTTCTCTGTTCAGCTAGAGAACTGCAAGCTGAGCGCTAC 1565  
QY 565 ATGCTATGCTCAGACGAAATGTTGAAGACGGGAGAGGCCCCAGATTTCAATGCCACCTTCTC 624  
DB 1564 TTGCAGTGGAGTCAACGGGTATAGGGAAGCATCTAAGGATTTTCATCCGAGAGCAGTTGCT 1505  
QY 625 CGG---CCCTATGTGGGCAATTCATCTGGCAATTGGCTCTGACTGGAAAGACGCTGTGCC 681  
DB 1504 CGGGGTGCTTTTGGGCAATTCATCTGGCAACCGTATCGATTGGGTGAGAGCCTGTGAG 1445  
QY 682 ATGCTGAAGGAGCGG-----GACTGCGAGCTCGCACTTTTCATGGCTCTCCCGAGGT 732  
DB 1444 CACGTCAAGGATAGCAGCATCTGTTGCCCTCGCGCAGTGTCTGGGCTATAAANAATAA 1385  
QY 733 GTGGGCTACAGCCGACGACAGCGGCCCTCTCAGATGACTATGTGCTCTGCTGACCTG 792  
DB 1384 CGTGGTGCACCTTACCCGAGCTCTGCAATGCCCTCCAAGGAGGCGATCATCCGCCAGCTA 1325  
QY 793 AAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGCCAGTCCGCTCTAC 852  
DB 1324 AAGAGAACCATTAAGAAGCTGCGCCAAACTCAGCCGGAACAAACGAATCAAAATCAGTTTC 1265  
QY 853 GTTGTCTACTGATTCGAGAGTTATGTGCTCTGAGCTTCCAAACAGCTTCTCAAAGGGAAGTG 912  
DB 1264 GTGGGTCAGACTCAATCAGATGATGTTGTTGAACTAAACAGCGGCTTTAGTGCATGGGC 1205  
QY 913 AAGGTGGTGGAGCTGAAAGCCCTGAGGTGG---CCAGGTGCGACCTGTATCATCTCGGCCAA 969  
DB 1204 ATCAGTGTGCAACAGCTGCCGAGGATGATCTTTACCTGGACTTGGCCATTTCTCGGACAG 1145  
QY 970 GCGACCACTTTATTTGGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1029  
DB 1144 TCGAACCACTTTATCGGCACTGTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1085  
QY 1030 GACCTCCAGGGGAGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1071  
DB 1084 GATGTGCAGGTTTTTCCATCTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1043

Search completed: November 9, 2005, 04:43:18  
Job time : 887.842 secs

THIS PAGE BLANK (USPTO)





Qy 61 TTTGGGAACCAAGCGCATCATCTTTGGGCTCTTGGCATTTGCAAAAGCTGCTAAACCGT 120  
Db TTTGGGAACCAAGCGCATCATCTTTGGGCTCTTGGCATTTGCAAAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGAGATCACAAAGCCTCTCTTTACCAACCTC 180  
Db ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGAGATCACAAAGCCTCTCTTTACCAACCTC 180  
Qy 181 CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGAGCTTACCATCGGCTCATC 240  
Db CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGAGCTTACCATCGGCTCATC 240  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGCCCCCTCGAAGCGGGTG 300  
Db AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGCCCCCTCGAAGCGGGTG 300  
Qy 301 GATATCTGCTTGGAGTGGAGCGACGAAAGCCAGATGAAGAACGCTGCCCCATGAAG 360  
Db GATATCTGCTTGGAGTGGAGCGACGAAAGCCAGATGAAGAACGCTGCCCCATGAAG 360  
Qy 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCCGAG 420  
Db GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCCGAG 420  
Qy 421 CTTTTTACAGGCATTTCTTCAAGTGTCTTACAGAGAACAAATGGAGCCAGAGATTTTCT 480  
Db CTTTTTACAGGCATTTCTTCAAGTGTCTTACAGAGAACAAATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGCTGTCTGCTGCCAGAGGCCCGCCAGTTCCCGCTCTAGAA 540  
Db CCAAGGAACATCCGCTGTCTGCTGCCAGAGGCCCGCCAGTTCCCGCTCTAGAA 540  
Qy 541 GAACACAGGCCTACAGAGTACATGTATGTGTCAGAGCAATGTGAGAGCGGAG 600  
Db GAACACAGGCCTACAGAGTACATGTATGTGTCAGAGCAATGTGAGAGCGGAG 600  
Qy 601 GCCAGATTCATGCCCATCTTGTGCGGCCCTATGTGGGCAATCATCTGGCAATGGCTCT 660  
Db GCCAGATTCATGCCCATCTTGTGCGGCCCTATGTGGGCAATCATCTGGCAATGGCTCT 660  
Qy 661 GACTGGAAGAACGCTGTGCCATGTGTAAGACGGGACTCGAGGCTCGCACTTCATGCCC 720  
Db GACTGGAAGAACGCTGTGCCATGTGTAAGACGGGACTCGAGGCTCGCACTTCATGCCC 720  
Qy 721 TCTCCGAGTGTGTGGCTACAGCCGACGACAGCGGCCCTCCACGATGACTATGTGC 780  
Db TCTCCGAGTGTGTGGCTACAGCCGACGACAGCGGCCCTCCACGATGACTATGTGC 780  
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGATGCC 840  
Db CTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGATGCC 840  
Qy 841 CAGTGGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCACAGCTCTTC 900  
Db CAGTGGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCACAGCTCTTC 900  
Qy 901 AAAGGAAGGTGAAGTGTGAGCTTGAAGCTGAGGTGGCCAGGTGACCTGTACATC 960  
Db AAAGGAAGGTGAAGTGTGAGCTTGAAGCTGAGGTGGCCAGGTGACCTGTACATC 960  
Qy 961 CTCGGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAG 1020  
Db CTCGGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAG 1020  
Qy 1021 CGGAGCGGGAACCTTCAGGGGAGGCGCTCTTCTTCTTGGCATGACAGGCCCTCTAAG 1080  
Db CGGAGCGGGAACCTTCAGGGGAGGCGCTCTTCTTCTTGGCATGACAGGCCCTCTAAG 1080  
Qy 1081 CTGCGGGAACGAGTTCTGATTTCTGGCGGAGACCAAGCCCTCTGTATCTTGGAGGACAG 1140  
Db CTGCGGGAACGAGTTCTGATTTCTGGCGGAGACCAAGCCCTCTGTATCTTGGAGGACAG 1140  
Qy 1141 AGTCTGAGCTGCTCTTCCAGGCGAGGCTGGGAGCGAGAGGTGCTCCGGGATTGCAAACT 1200

Db 1141 AGTCTGAGCTGCTCTTCCAGCCAGGCTTGGCAGCAGAGGTGCTCCGGGATTGCAAACT 1200  
Qy 1201 CTTCTTTCTACCTGSCAAAGATGGAGAGAGTCCAGGACCCCTCAAGGAGGAGAGCGC 1260  
Db CTTCTTTCTACCTGSCAAAGATGGAGAGAGTCCAGGACCCCTCAAGGAGGAGAGCGC 1260  
Qy 1261 TCCATATCCAGGSCATAGGACTTGCAGGTTCTTAGAGCAGGAGCATCTCCATCGCAC 1320  
Db TCCATATCCAGGSCATAGGACTTGCAGGTTCTTAGAGCAGGAGCATCTCCATCGCAC 1320  
Qy 1321 GTGCTTCTGCTCTTCTGGGAATTTCTACACTGGCAGAAAGAGTCCAGGCTCCGCTTTCT 1380  
Db GTGCTTCTGCTCTTCTGGGAATTTCTACACTGGCAGAAAGAGTCCAGGCTCCGCTTTCT 1380  
Qy 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTCTTCAAGAGAGATTTTTTATAGA 1440  
Db GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTCTTCAAGAGAGATTTTTTATAGA 1440  
Qy 1441 GAGATTTCTATAATTTTTGATACAAAGTCAAGGTCATGACTATCTAGAACTCTCTGTGTTTTGA 1500  
Db GAGATTTCTATAATTTTTGATACAAAGTCAAGGTCATGACTATCTAGAACTCTCTGTGTTTTGA 1500  
Qy 1501 AAATCATTGAATTC 1514  
Db 1501 AAATCATTGAATTC 1514

## RESULT 2

US-09-333-729A-2  
; Sequence 2, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P1041PDI-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 2  
; LENGTH: 1514  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-333-729A-2

Query Match 100.0%; Score 1514; DB 3; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCGCGGGCTCCTGGGACCGCGCGGTACCTGCTCTACTGCCCTGCATGGGCGC 60  
Db 1 ATGCCGCGGGCTCCTGGGACCGCGCGGTACCTGCTCTACTGCCCTGCATGGGCGC 60  
Qy 61 TTTGGGAACCAAGCGCATCATCTTTGGGCTCTCTGGGATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACCAAGCGCATCATCTTTGGGCTCTCTGGGATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGAGCTCAAGGCTCTTTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGAGCTCAAGGCTCTTTTACCAACCTC 180  
Qy 181 CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 181 CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGGCCCTCGAAGCGGGTG 300  
Db 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGGCCCTCGAAGCGGGTG 300  
Qy 301 GCATACTCTTTGAGGTGGCAGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

Db 301 GCATACCTCTTTAGGTGGCCAGCCAGCGAAGCCAGATAAGAGACGTGCCCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTACAGGCAATTCCTTCAGTGCTTCCACAGAGAAATGAGGACGAGATTTTCT 480  
Db 421 CTTTTACAGGCAATTCCTTCAGTGCTTCCACAGAGAAATGAGGACGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGCTGCTTCCCTGCGAGAGCCCGAGGCTTCCCGTCTAGAA 540  
Db 481 CCAAGGAACATCCGCTGCTTCCCTGCGAGAGCCCGAGGCTTCCCGTCTAGAA 540  
Qy 541 GAACACGCGCACTACAGAGTACATGATGTGCTCAGACGAATGCTCAAGACGGAGAG 600  
Db 541 GAACACGCGCACTACAGAGTACATGATGTGCTCAGACGAATGCTCAAGACGGAGAG 600  
Qy 601 GCCCAGATTCATGCCACCTTTGTCGGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 660  
Db 601 GCCCAGATTCATGCCACCTTTGTCGGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 660  
Qy 661 GACTGGAAGACGCTGTGCCATGCTGAAGACGGGACTGCAAGGCTGCATTTGAGGCC 720  
Db 661 GACTGGAAGACGCTGTGCCATGCTGAAGACGGGACTGCAAGGCTGCATTTGAGGCC 720  
Qy 721 TCTCGCAGTGTGGGCTACAGCGCAGCAGCGGCCCGCCCTCAGCATCATATGTC 780  
Db 721 TCTCGCAGTGTGGGCTACAGCGCAGCAGCGGCCCGCCCTCAGCATCATATGTC 780  
Qy 781 CTGCTGACCTGAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 840  
Db 781 CTGCTGACCTGAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 840  
Qy 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTCTGAGCTCCAAAGCTTTTC 900  
Db 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTCTGAGCTCCAAAGCTTTTC 900  
Qy 901 AAAGGGAAGTGAAGTGTGCTGAGCTGAAGCTGAGGTGGCCAGGTCGACTCTACATC 960  
Db 901 AAAGGGAAGTGAAGTGTGAGCTGAAGCTGAGGTGGCCAGGTCGACTCTACATC 960  
Qy 961 CTCGCCCAAGCGGACCACTTTATTGGCACTGTGTCTCTCTTCTTCTGAGTGGAG 1020  
Db 961 CTCGCCCAAGCGGACCACTTTATTGGCACTGTGTCTCTCTTCTTCTGAGTGGAG 1020  
Qy 1021 CGGAGCGGGACCTCCAGGGAGGCGCTTCTTTTTCGGCATGGACAGGCCCTTAAAG 1080  
Db 1021 CGGAGCGGGACCTCCAGGGAGGCGCTTCTTTTTCGGCATGGACAGGCCCTTAAAG 1080  
Qy 1081 CTGCGGACGAGTCTTGATTTCTGCGGAGCAGCAGACCTCTGATCTGAGGAGCCAG 1140  
Db 1081 CTGCGGACGAGTCTTGATTTCTGCGGAGCAGCAGACCTCTGATCTGAGGAGCCAG 1140  
Qy 1141 AGTCTGAGTCTGCTTCCAGCGAGGCTGCGAGCAGAGTGTCTCGGATTCGAACT 1200  
Db 1141 AGTCTGAGTCTGCTTCCAGCGAGGCTGCGAGCAGAGTGTCTCGGATTCGAACT 1200  
Qy 1201 CCTCTTCTACCTGCAAGATGAGAGAGTGCAGGACCTTCAAGGAGGAGAGCGC 1260  
Db 1201 CCTCTTCTACCTGCAAGATGAGAGAGTGCAGGACCTTCAAGGAGGAGAGCGC 1260  
Qy 1261 TCCATATCCAGGCGATAGGATTCAGGTTCTTAGGAGGAGGATCTCCCATCGCAC 1320  
Db 1261 TCCATATCCAGGCGATAGGATTCAGGTTCTTAGGAGGAGGATCTCCCATCGCAC 1320  
Qy 1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCTTCTTCT 1380  
Db 1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCTTCTTCT 1380  
Qy 1381 GGTCCACTCTGCTGAGCAGCCTGGGATGCTGAATCTTCCAGAGAGATTTTATAGA 1440

Db 1381 GGTCCACTCTGCTGAGCAGCCTGGGATGCTGAATCTTCCAGAGAGATTTTATAGA 1440  
Qy 1441 GAGATTTCTATAATTTTATACAGAGTTCATGACTATCTCTGCTGCTGTTTGA 1500  
Db 1441 GAGATTTCTATAATTTTATACAGAGTTCATGACTATCTCTGCTGCTGTTTGA 1500  
Qy 1501 AAATCAATTGAATTC 1514  
Db 1501 AAATCAATTGAATTC 1514

## RESULT 3

US-08-978-741-5  
; Sequence 5, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978.741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11284 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-978-741-5

Query Match 99.9%; Score 1512.4; DB 3; Length 11284;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGCTCTCGGACCGCGCGGTTCCTGCTCTACTGCTCTGATCGGGCGC 60  
Db 4236 ATGCCCGCGGCTCTCGGACCGCGCGGTTCCTGCTCTACTGCTCTGATCGGGCGC 4295  
Qy 61 TTGCGGAACAGCGCGCATCTTCTGCGCTCTCTGCGATTTGCAAGCTCTAAACCGT 120  
Db 4296 TTGCGGAACAGCGCGCATCTTCTGCGCTCTCTGCGATTTGCAAGCTCTAAACCGT 4355  
Qy 121 ACCTTGGCTGTCCCTCTCTGAGTACAGCATCAGAGCTCTCTTCCACCAACCTC 180  
Db 4356 ACCTTGGCTGTCCCTCTCTGAGTACAGCATCAGAGCTCTCTTCCACCAACCTC 4415  
Qy 181 CATGTGCTCTACAGAGTACTTCAAGCTGAGGCCCTCAGGCTTACCATCGGGTCTATC 240  
Db 4416 CATGTGCTCTACAGAGTACTTCAAGCTGAGGCCCTCAGGCTTACCATCGGGTCTATC 4475

Qy 241 AGCTTGAGGATTTTCATGAGAAAGCTGGCACCACCCACTGGCCCCCTCGAGAAGCGGGT 300  
Db 4476 AGCTTGAGGATTTTCATGAGAAAGCTGGCACCACCCACTGGCCCCCTCGAGAAGCGGGT 4535  
Qy 301 GCATACCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGCCCCATGAAG 360  
Db 4536 GCATACCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGCCCCATGAAG 4595  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 4596 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTCATGTGAGTTTCAACAAGTCGGAG 4655  
Qy 421 CTTTATTAGAGCATTTCTCTGAGTCTCTCTACAGAGCAATGAGCAGAGATTTTCT 480  
Db 4656 CTTTATTAGAGCATTTCTCTGAGTCTCTCTACAGAGCAATGAGCAGAGATTTTCT 4715  
Qy 481 CCAAGGAACATCCGGTGTCCCTGCGAGAGCCCGCCAGTTCCTCGTCTTAGAA 540  
Db 4716 CCAAGGAACATCCGGTGTCCCTGCGAGAGCCCGCCAGTTCCTCGTCTTAGAG 4775  
Qy 541 GAACACAGGCCACTACAGAAATGATGTTGTCAGACGAAATGTTGAAGACGGGAGAG 600  
Db 4776 GAACACAGGCCACTACAGAAATGATGTTGTCAGACGAAATGTTGAAGACGGGAGAG 4835  
Qy 601 GCCAGATTCATGCCACCTTTGCGGCCCTTATGTGGCAATTCATCTGCGAATGGCTCT 660  
Db 4836 GCCAGATTCATGCCACCTTTGCGGCCCTTATGTGGCAATTCATCTGCGAATGGCTCT 4895  
Qy 661 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACTGCGAGGCTCGCACTTCATGCC 720  
Db 4896 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACTGCGAGGCTCGCACTTCATGCC 4955  
Qy 721 TCTCCGAGTGTGGGTACAGCGCAGCACAGCGCCCCCTCACGATGACTATGTC 780  
Db 4956 TCTCCGAGTGTGGGTACAGCGCAGCACAGCGCCCCCTCACGATGACTATGTC 5015  
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCCGTCGATGCC 840  
Db 5016 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCCGTCGATGCC 5075  
Qy 841 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGTCCCAAGCTCTTC 900  
Db 5076 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGTCCCAAGCTCTTC 5135  
Qy 901 AAAGGAAGGTGAAGGTGGTGAGCCTGAAGCTGAGTGGCCAGGTGCACTGTACATC 960  
Db 5136 AAAGGAAGGTGAAGGTGGTGAGCCTGAAGCTGAGTGGCCAGGTGCACTGTACATC 5195  
Qy 961 CTCGCCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTTCACTGCCCTTTGTGAAG 1020  
Db 5196 CTCGCCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTTCACTGCCCTTTGTGAAG 5255  
Qy 1021 CGGAGCGGACCTCCAGGGGAGCGCTCTTCTTCTCGGCATGACAGGCCCTTAAG 1080  
Db 5256 CGGAGCGGACCTCCAGGGGAGCGCTCTTCTTCTCGGCATGACAGGCCCTTAAG 5315  
Qy 1081 CTGCGGACGAGTTCGATTCGCGCGGAGCACCAGACCTCTGATCCTGGAGGAGCCAG 1140  
Db 5316 CTGCGGACGAGTTCGATTCGCGCGGAGCACCAGACCTCTGATCCTGGAGGAGCCAG 5375  
Qy 1141 AGTCTGAGCTGGCTTCTCAGCAGGCTTGGAGCGGAGTGTCTCCGGATTCGAACT 1200  
Db 5376 AGTCTGAGCTGGCTTCTCAGCAGGCTTGGAGCGGAGTGTCTCCGGATTCGAACT 5435  
Qy 1201 CCTCTTCTCACTGCCAAAGTGAAGAGTGCAGGGAACCCCTCAAGAGGGAGAGCGC 1260  
Db 5436 CCTCTTCTCACTGCCAAAGTGAAGAGTGCAGGGAACCCCTCAAGAGGGAGAGCGC 5495  
Qy 1261 TCCATATCCAGGCGATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
Db 5496 TCCATATCCAGGCGATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 5555  
Qy 1321 GTGCTTTCTGCTCTCTCTGGGAATTTCTCACATGGCAAGCAGTCCAGCTCCGCTCTTCT 1380

Db 5556 GTGCTTTCTGCTCTCTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTCTTCT 5615  
Qy 1381 GGTCCACTCTGCTCTCTGAGCAGCCTGGGATGCTGAACCTTTTCAGAGAGATTTTTTATAGA 1440  
Db 5616 GGTCCACTCTGCTCTCTGAGCAGCCTGGGATGCTGAACCTTTTCAGAGAGATTTTTTATAGA 5675  
Qy 1441 GAGATTTCTATAATTTTGTATCAAGGTCATGACTATCTCTAGAACTCTCTGTGGTTTTTGA 1500  
Db 5676 GAGATTTCTATAATTTTGTATCAAGGTCATGACTATCTCTAGAACTCTCTGTGGTTTTTGA 5735  
Qy 1501 AAATCATTTGAATTC 1514  
Db 5736 AAATCATTTGAATTC 5749

RESULT 4  
US-08-978-741-7  
; Sequence 7, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Pucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5009 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; US-08-978-741-7

Query Match 95.68; Score 1447.4; DB 3; Length 5009;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 1448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 66 GAACAGAGCCGATCAGTCTTGGGCTCTGCGCATTTGCAAGCTCTAAACCGTACCTT 125  
Db 1 GAACAGAGCCGATCAGTCTTGGGCTCTTGGGCTCTTGGCATTTGCAAGCTCTAAACCGTACCTT 60  
Qy 126 GGTCTGCTCCTCTTGGATTGAGTACCAGCATCAAGGCTCCTTTACCAAGCTTCATGT 185  
Db 61 GGTCTGCTCCTCTTGGATTGAGTACCAGCATCAAGGCTCCTTTACCAAGCTTCATGT 120  
Qy 186 GTCTTACCAGAGTACTTTCAAGCTGGAGCCCTTCAGGCTTACATCGGGTTCATGCTT 245  
Db 121 GTCTTACCAGAGTACTTTCAAGCTGGAGCCCTTCAGGCTTACATCGGGTTCATGCTT 180



Qy 546 CAGGCCACTACAACTACATGTTATGTCAGACGAAATGGTGAAGACGGAGAGGCCCA 605  
Db 481 CAGGCCACTACAACTACATGTTATGTCAGACGAAATGGTGAAGACGGAGAGGCCCA 540  
Qy 606 GATTATGCTCCCACTTTGTCGGCCCTATGTTGGGCATTCATCTGCGCATTTGGCTCTGACTG 665  
Db 541 GATTATGCTCCCACTTTGTCGGCCCTATGTTGGGCATTCATCTGCGCATTTGGCTCTGACTG 600  
Qy 666 GAAGAACGCTGTGCTCATCTCAAGACGGGACTGACAGGCTGCACATTCATGGCTCTCC 725  
Db 601 GAAGAACGCTGTGCTCATCTCAAGACGGGACTGACAGGCTGCACATTCATGGCTCTCC 660  
Qy 726 GCAGTGTGTGGCTACAGCCGACAGACAGCGGCCCTCTCAGATGACTATGTGCTGCC 785  
Db 661 GCAGTGTGTGGCTACAGCCGACAGACAGCGGCCCTCTCAGATGACTATGTGCTGCC 720  
Qy 786 TGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGCCTGGATGCCCAATC 845  
Db 721 TGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGCCTGGATGCCCAATC 780  
Qy 846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTCAAAGG 905  
Db 781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTCAAAGG 840  
Qy 906 GAAAGTGAAGTGGTGAAGCTCAAGCCTGAGGTGGCCAGGTGACCTGTATCATCTCGG 965  
Db 841 GAAAGTGAAGTGGTGAAGCTCAAGCCTGAGGTGGCCAGGTGACCTGTATCATCTCGG 900  
Qy 966 CAAAGCCGACACTTTATTTGGCACTGTGCTCTCTCTCTCACTGCTCTTTGTGAAGCGGGA 1025  
Db 901 CAAAGCCGACACTTTATTTGGCACTGTGCTCTCTCTCTCACTGCTCTTTGTGAAGCGGGA 960  
Qy 1026 GCGGAGCTCCAGGAGGAGCCGTCTTTCTTCTCGGATGAGCAGGCCCCCTTAAGCTGCG 1085  
Db 961 GCGGAGCTCCAGGAGGAGCCGTCTTTCTTCTCGGATGAGCAGGCCCCCTTAAGCTGCG 1020  
Qy 1086 GCAGAGTCTCTGATTTCTGCGGAGCACCAGACCTCTGATCTGAGGAGGACAGAGTCT 1145  
Db 1021 GCAGAGTCTGATTTCTGCGGAGCACCAGACCTCTGATCTGAGGAGGACAGAGTCT 1080  
Qy 1146 GAGCTGTCTCTCCAGCAGGCTGGAGCCAGAGGTGCTCGGATTTGCAACTCTCTCT 1205  
Db 1081 GAGCTGTCTCTCCAGCAGGCTGGAGCCAGAGGTGCTCGGATTTGCAACTCTCTCT 1140  
Qy 1206 TCTCACTCTGCAAGATGGAGAGAGTGCAGAGGACCCCTCAAGAGGAGGAGCGCTCCAT 1265  
Db 1141 TCTCACTCTGCAAGATGGAGAGAGTGCAGAGGACCCCTCAAGAGGAGGAGCGCTCCAT 1200  
Qy 1266 ATCCAGGCGATAGGACTTTGCAAGTTCCTAGGAGCAGGACATCTCCATGCGACGTGCT 1325  
Db 1201 ATCCAGGCGATAGGACTTTGCAAGTTCCTAGGAGCAGGACATCTCCATGCGACGTGCT 1260  
Qy 1326 TTCTGCTCTTCTGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCTCTCTGCTCC 1385  
Db 1261 TTCTGCTCTTCTGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCTCTCTGCTCC 1320  
Qy 1386 ACTGCTCTGAGGAGCTGGAGTCTGAACTCTTCAAGAGATTTTTTATAGAGAGAT 1445  
Db 1321 ACTGCTCTGAGGAGCTGGAGTCTGAACTCTTCAAGAGATTTTTTATAGAGAGAT 1380  
Qy 1446 TTCTATAATTTTATACAAAGTCTGACTATCTCTAGAACTCTCTGCTGCTTTTGAATC 1505  
Db 1381 TTCTATAATTTTATACAAAGTCTGACTATCTCTAGAACTCTCTGCTGCTTTTGAATC 1440  
Qy 1506 ATTGAATTC 1514  
Db 1441 ATTGAATTC 1449

RESULT 6  
US-08-978-741-4  
; Sequence 4, Application US/08978741  
; Patent No. 6100076

GENERAL INFORMATION:  
APPLICANT: Yang Wang, Michael W. Spellman  
TITLE OF INVENTION: O-Fucosyltransferase  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,741  
FILING DATE: 26-No. 6100076-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/792498  
FILING DATE: 31  
ATTORNEY/AGENT INFORMATION:  
NAME: Svoboda, Craig G.  
REGISTRATION NUMBER: 39,044  
REFERENCE/DOCKET NUMBER: P1041P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1489  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1300 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-978-741-4  
Query Match 76.8%; Score 1163.4; DB 3; Length 1300;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCGCGGCTCTCGGACCGGCGGTTACTGCTCTACTGCCCCGATGGGGCGC 60  
Db 136 ATGCCCGCGGCTCTCGGACCGGCGGTTACTGCTCTACTGCCCCGATGGGGCGC 195  
Qy 61 TTGGGAAACAGGCGCATCATCTTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT 120  
Db 196 TTGGGAAACAGGCGCATCATCTTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT 255  
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCAAGGCTCTTTTACCACCACTC 180  
Db 256 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCAAGGCTCTTTTACCACCACTC 315  
Qy 181 CATGTGCTTACAGAGTACTTCAAGCTGGAGGCCCTCCAGGCTTACCATCGGTCATC 240  
Db 316 CATGTGCTTACAGAGTACTTCAAGCTGGAGGCCCTCCAGGCTTACCATCGGTCATC 375  
Qy 241 AGCTTGGAGGATTTTATGAGAGAGCTGGGACCCACCCACTGGGCCCTCGAAGCGGGTG 300  
Db 376 AGCTTGGAGGATTTTATGAGAGAGCTGGGACCCACCCACTGGGCCCTCGAAGCGGGTG 435  
Qy 301 GCATACCTCTTTGAGGTGGCAGCCAGGAGGCCAGATAAGAGACGTCGCCCATGAAG 360  
Db 436 GCATACCTCTTTGAGGTGGCAGCCAGGAGGCCAGATAAGAGACGTCGCCCATGAAG 495  
Qy 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 496 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 555  
Qy 421 CTTTATTACAGGCAATTTCTTCACTGCTCTTCTTACAGAGCAATGGAGCCAGAGATTTTCT 480  
Db 556 CTTTATTACAGGCAATTTCTTCACTGCTCTTCTTACAGAGCAATGGAGCCAGAGATTTTCT 615



Db 1156 CGGAGCGGACCTCCAGGGAGCGGCTCTCTTCTTCGGCATGACAGGCCCCCTAAG 1215  
Qy 1081 CTGCGGAGCAGATTCTGATTTCTGCGCGGAGCACAGACCTCTGTATCTTGAAGGAGCCAG 1140  
Db 1216 CTGCGGAGCAGATTCTGATTTCTGCGCGGAGCACAGACCTCTGTATCTTGAAGGAGCCAG 1275  
Qy 1141 AGTCTGAGCTGGTCTCTTCCAGCCAG 1165  
Db 1276 AGTCTGAGCTGGTCTCTTCCAGCCAG 1300

## RESULT 8

US-08-978-741-16  
; Sequence 16, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1100 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-978-741-16

Query Match 72.7%; Score 1100; DB 3; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGCCCCGGGGCTCTGGGACCGGCGGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Db 1 ATGCCCCGGGGCTCTGGGACCGGCGGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Qy 61 TTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGCATTTTGCATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGCATTTTGCATTTTGCATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCTCTCTGGATTGAGTACCAAGCTCCTTTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCTCTCTGGATTGAGTACCAAGCTCCTTTTACCAACCTC 180  
Qy 181 CATGTGCTTACCAAGTACTTCAAGTGGAGCCCTCCAGGCTTACATCGGTGATC 240  
Db 181 CATGTGCTTACCAAGTACTTCAAGTGGAGCCCTCCAGGCTTACATCGGTGATC 240

Qy 241 AGCTTGGAGGATTTCTATGGAGAGCTGGGACCCACCCACTGGCCCCCTCGAAGCGGGTG 300  
Db 241 AGCTTGGAGGATTTCTATGGAGAGCTGGGACCCACCCACTGGCCCCCTCGAAGCGGGTG 300  
Qy 301 GCATACTCTTCTTGAAGTGGCAGCCAGGAGGAGCCAGATAAGAGACGCTGCCCCCATGAAG 360  
Db 301 GCATACTCTTCTTGAAGTGGCAGCCAGGAGGAGCCAGATAAGAGACGCTGCCCCCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCTATGTGAGTTTCAACAGTCCGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCTATGTGAGTTTCAACAGTCCGAG 420  
Qy 421 CTTTATTACAGGCATTTCTTCTAGTGTCTTCTACAGAGAACATGGAGCCAGAGATTTCT 480  
Db 421 CTTTATTACAGGCATTTCTTCTAGTGTCTTCTACAGAGAACATGGAGCCAGAGATTTCT 480  
Qy 481 CCAAAGGAACATCCGGTGTCTGCCCTGGCAGGAGCCCGAGCCAGCTTCCCGTCCCTAGAA 540  
Db 481 CCAAAGGAACATCCGGTGTCTGCCCTGGCAGGAGCCCGAGCCAGCTTCCCGTCCCTAGAA 540  
Qy 541 GAAACACAGGCCACTTACAGAAATGATGTATGGTTCAGACGAAATGGTGAAGACGGGAG 600  
Db 541 GAAACACAGGCCACTTACAGAAATGATGTATGGTTCAGACGAAATGGTGAAGACGGGAG 600  
Qy 601 GCCCAGATTCTATGCCCATCTTGTTCGGCCCTATGTGGGCAATTCATCTCGCATTTGGCTCT 660  
Db 601 GCCCAGATTCTATGCCCATCTTGTTCGGCCCTATGTGGGCAATTCATCTCGCATTTGGCTCT 660  
Qy 661 GACTGGAAGAACCGCTGTGCGCATGTGCGATGCGGAGCGGAGCTGCGAGCTTCATGCGCC 720  
Db 661 GACTGGAAGAACCGCTGTGCGCATGTGCGATGCGGAGCGGAGCTGCGAGCTTCATGCGCC 720  
Qy 721 TCTCCGCAAGTGTGGGCTACAGCGCGAGCAGCGGCCCTCCACGATGACTATGTGC 780  
Db 721 TCTCCGCAAGTGTGGGCTACAGCGCGAGCAGCGGCCCTCCACGATGACTATGTGC 780  
Qy 781 CTGCTGACCTTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
Db 781 CTGCTGACCTTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
Qy 841 CAGTGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTTCAACAGCTCTTC 900  
Db 841 CAGTGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTTCAACAGCTCTTC 900  
Qy 901 AAAGGGAAGGTGAAGGTGGTGAAGCTTGAAGCTGAGGTGGCCCGAGCTGACCTGTACATC 960  
Db 901 AAAGGGAAGGTGAAGGTGGTGAAGCTTGAAGCTGAGGTGGCCCGAGCTGACCTGTACATC 960  
Qy 961 CTCGGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 961 CTCGGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Qy 1021 CGGAGCGGAGCTCCAGGGAGGCGCTCTTCTTCTTTCGGCATGACAGGCCCTTAAAG 1080  
Db 1021 CGGAGCGGAGCTCCAGGGAGGCGCTCTTCTTCTTTCGGCATGACAGGCCCTTAAAG 1080  
Qy 1081 CTGCGGAGCAGATTCTGATT 1100  
Db 1081 CTGCGGAGCAGATTCTGATT 1100

## RESULT 9

US-09-333-729A-16  
; Sequence 16, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P1041P1D1-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741



```

; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 16
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-16

```

```
Query Match          72.7%; Score 1100; DB 3; Length 1100;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

|    |     |   |     |
|----|-----|---|-----|
| Qy | 1   | ATGCCGCGGGCTCCTTGGGACCCGGCCGGTTTACCTGCTCTACTGCCCCCTGCATGGGGGCGC | 60  |
| Db | 1   | ATGCCGCGGGCTCCTTGGGACCCGGCCGGTTTACCTGCTCTACTGCCCCCTGCATGGGGGCGC | 60  |
| Qy | 61  | TTTGGGAACACGAGCGCGATCAGTCTCTGGGCTCTCGGCATTTGCAAGCTGCTAAACCGT    | 120 |
| Db | 61  | TTTGGGAACACGAGCGCGATCAGTCTCTGGGCTCTCGGCATTTGCAAGCTGCTAAACCGT    | 120 |
| Qy | 121 | ACCTTGGCTGTCCCTCTTGGATTGAGTACCAGCATCACAGCCTCTCTTTCAACCAACCTC    | 180 |
| Db | 121 | ACCTTGGCTGTCCCTCTTGGATTGAGTACCAGCATCACAGCCTCTCTTTCAACCAACCTC    | 180 |
| Qy | 181 | CATGTGCTTACAGAAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGGTCATC       | 240 |
| Db | 181 | CATGTGCTTACAGAAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGGTCATC       | 240 |
| Qy | 241 | AGCTTGAGGATTTTCATGGAGAGCTGGCACCACCCACCTGGCCCCCTGAGAAAGGGGTG     | 300 |
| Db | 241 | AGCTTGAGGATTTTCATGGAGAGCTGGCACCACCCACCTGGCCCCCTGAGAAAGGGGTG     | 300 |
| Qy | 301 | GCATACTGCTTTGAGTGGGACGCCAGCAAGCCACAGATAAGAAGACGTGCCCATGAAG      | 360 |
| Db | 301 | GCATACTGCTTTGAGTGGGACGCCAGCAAGCCACAGATAAGAAGACGTGCCCATGAAG      | 360 |
| Qy | 361 | GAAGGAAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG     | 420 |
| Db | 361 | GAAGGAAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG     | 420 |
| Qy | 421 | CTTTTTCACAGGCATTTCTTTCAGTGCTTCTTACAGAGAACATGAGGACACAGATTTTCT    | 480 |
| Db | 421 | CTTTTTCACAGGCATTTCTTTCAGTGCTTCTTACAGAGAACATGAGGACACAGATTTTCT    | 480 |
| Qy | 481 | CCAAAGGAAACATCCGGTGCTTGGCCCTGCGCAGGAGCCCGACCCAGTTTCCCGCTCTAGAA  | 540 |
| Db | 481 | CCAAAGGAAACATCCGGTGCTTGGCCCTGCGCAGGAGCCCGACCCAGTTTCCCGCTCTAGAA  | 540 |
| Qy | 541 | GAACACAGGCCACTACAGAAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAG     | 600 |
| Db | 541 | GAACACAGGCCACTACAGAAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAG     | 600 |
| Qy | 601 | GCCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTGCCGATTTGGCTCT     | 660 |
| Db | 601 | GCCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTGCCGATTTGGCTCT     | 660 |
| Qy | 661 | GACTTGAAGAACGCTCTGTGCATGCTGAAGACGGGACTGCAGGCTCGCACTTCATGCCC     | 720 |
| Db | 661 | GACTTGAAGAACGCTCTGTGCATGCTGAAGACGGGACTGCAGGCTCGCACTTCATGCCC     | 720 |
| Qy | 721 | TCTCCGAGTGTGTGGGCTACAGCCGAGCACAGCGGCCCCCTTCACGATGACTATGTGC      | 780 |
| Db | 721 | TCTCCGAGTGTGTGGGCTACAGCCGAGCACAGCGGCCCCCTTCACGATGACTATGTGC      | 780 |
| Qy | 781 | CTGGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC     | 840 |
| Db | 781 | CTGGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC     | 840 |
| Qy | 841 | CAGTTCGGTCTACGTTGCTACTGATTCGAGAGTTTATGTGCTGAGCTCCACACAGCTCTTC   | 900 |
| Db | 841 | CAGTTCGGTCTACGTTGCTACTGATTCGAGAGTTTATGTGCTGAGCTCCACACAGCTCTTC   | 900 |
| Qy | 901 | AAAGGGAAGGTGAAGTGGTGAGCCTGAAGCCTTGAGGTGGGCCACAGGTCGACCTGTACATC  | 960 |

## RESULT 10

```

US-09-270-767-14353
;
; Sequence 14353, Application US/09270767
; Patent No. 6703491
;
; GENERAL INFORMATION:
;
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14353

```

| Query Match.          | 11.1%  | Score 168;  | DB 4;           | Length 1320;       |
|-----------------------|--------|---|-----------------|--------------------|
| Best Local Similarity | 51.6%; | Pred. No. 6.1e-41;  |                 |                    |
| Matches               | 564;   | Conservative 0;   | Mismatches 485; | Indels 45; Gaps 6; |
| Qy                    | 17     | GGGACCGCGCGGTACCTGCTCTACTGCCCCCTGCATGGGCGCTTTGGGAACACGAGCCG     | 76              |                    |
| Db                    | 73     | GCATCCCAATGGCTACCTACCTACTGTCGCTGATGGACGCTTTGGCAACACGAGCCG       | 132             |                    |
| Qy                    | 77     | ATCACTTTCTGGGCTCTCTGGCATTTTGCAAGCTGTAAACCGTACCTTGGCTGTCCCTC     | 136             |                    |
| Db                    | 133    | ACCACCTCTCTGGGATCATTTGGCCTCTGCCAAGCGCTTAATCGACCCCTGATCCTGCCG    | 192             |                    |
| Qy                    | 137    | CTTGGATTGATPACAGCATCAACAAGCTCTTTTACCAACCTTCATGTCTCTACCA         | 196             |                    |
| Db                    | 193    | CGTGGGTGGAGT-----ATCGTAGGGGTGAACGTGGATCCCGCTCAGGTACCGTTCAACA    | 246             |                    |
| Qy                    | 197    | AGTACTTCAAGCTGGAGCCCCCTCCAGSGCTTACCATCGGGTCATCAGCTTGGAGGATTTCA  | 256             |                    |
| Db                    | 247    | CATATTTTGAATGGAGCCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCA    | 306             |                    |
| Qy                    | 257    | TGGAGAAGCTGGCACCCACCCACTGGCCCCCTCGAGAAGCGGGTGGCATACTGCTTTGAGG   | 316             |                    |
| Db                    | 307    | TGTGGCACCTGGCGGAGGACATTTGGCCAGATTCGGAGCGAGTGTCTATTTGTCTACAAG    | 366             |                    |
| Qy                    | 317    | -----TGGCAGCCACGAGCCAGCATAGAAGAAGCTGCGCCCCCATGAAGG              | 361             |                    |
| Db                    | 367    | AACGATATAGCCTTCACGAGGAGAGAAGCATCCAGACAAGCCCAATTCACGCGCAAGG      | 426             |                    |
| Qy                    | 362    | AAGGAACCCCTTTGGCCCCATCTCGGATCAGTTTTCATGTAGTTTCAACAAGTCGGAGC     | 421             |                    |
| Db                    | 427    | ATGGCAATCCTTTTGGTGCCTTTTGGGACACTTTTTCACATAGACTTTTGTGCGGTCTAGAGT | 486             |                    |
| Qy                    | 422    | TTTTTTCAGGCAATTCCTTTCAGTGTCTCTACAG-----AGAACATGTAGGCGAGA        | 472             |                    |
| Db                    | 487    | TCTATGGGCCATTCATTTTGTATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCA    | 546             |                    |
| Qy                    | 473    | GATTTTCTCCAAAGGAACATCCGGTGTCTGCCCTGCCAGAGCCCCAGCCAGTTCCTCCG     | 532             |                    |
| Db                    | 547    | AATATCCTCGAGAATCATATCCCGTACTCCGGTTTCAACCGAGCTCCGGGTAGTTTTCCTG   | 606             |                    |

|    |      |  |      |
|----|------|--|------|
| QY | 533  | TCCTAGAGAGACACAGGCCACTCAGAAAGTATCATGGTATGCTCAGACGAAATGGTGAAGA  | 592  |
| Db | 607  | TTTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTCAGTGCAGTACCAAGTATACGGGAAG  | 666  |
| QY | 593  | CGGAGAGAGCCCAAGATTCAATCCCAACCTTGTCCGGC---CCTATGTGGCAATTCATCTGC | 649  |
| Db | 667  | CATCTAAGGATTTCACTCCGAGAGCAGTTGCTCGGGTGCCCTTTTGGGCATTCATCTGC    | 726  |
| QY | 650  | GCATTGGCTCTGACTGGAGAACCCCTGTGCCATGCTGAAGACGG-----GACTG         | 700  |
| Db | 727  | GCAACGGTATTCGATTTGGGTGAGACCTGTGAGCAGCTCAGGATAGCCAGCACTCTGTTTG  | 786  |
| QY | 701  | CAGGCTGCGCACTTCATGSCCTCTCCGCAAGTGTGTGGGCTACAGCCGCAAGCAGCGGCC   | 760  |
| Db | 787  | CCTCGCGCAGCTGTGGGCTATAAAAATGAACGGTGGTGCACCTCTACCCGAGACTCTGCA   | 846  |
| QY | 761  | CCCTACAGATGACTATGTGCTGCTGCTGACTGTAAGAGAGATCCAGAGGGCTGTGAAGCTCT | 820  |
| Db | 847  | TGCGCTCCAAGGCGGATCATCGCCAGCTAAGAGAACCATTAAGAACGTGCCCAA         | 906  |
| QY | 821  | GGGTAGGTGCGTGATGCCCAGTCGGTCTAGTGTGCTACTGATTCGAGAGATTATGTGC     | 880  |
| Db | 907  | CTCAGCCGACAAACGAAATCAATTCAGTTTCGTGGCGTCAGACTCCAATCAATGATTTG    | 966  |
| QY | 881  | CTGAGCTCCAAACAGCTCTTCAAAGGGAAAG---TGAAGGTGTCAGCTGAAAGCCTGAGG   | 937  |
| Db | 967  | GTGAACCTAAACACGGCCCTTAGTCGATGGGCATCAGTGTGCAACAAGCTGACGAGGATG   | 1026 |
| QY | 938  | TGGCCAGGTTCGACCTGTACATCTCGCCCAAGCCGACCACTTTATTTGGCAACTGTGTCT   | 997  |
| Db | 1027 | ATCTTACCTTGGACTTGGCCATCTCGGACAGTTCGACCACTTTATCGGCACATGATAT     | 1086 |
| QY | 998  | CCTCTTCTACTGCTTTGTGAAGCGGAGCGGGAACCTCCAGGGAGGCCGTCTCTTTCT      | 1057 |
| Db | 1087 | CCTCTTACTCGGCATTCGTAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCT   | 1146 |
| QY | 1058 | TCGGCATGGACAGG   | 1071 |
| Db | 1147 | GGGGATTCCCCAAG   | 1160 |

RESULT 11

US-09-270-767-30507

; Sequence 30507, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 30507

; LENGTH: 722

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-30507

Db      228    CACTATACCCGAGGTCTGTGCATCGCCTTCCAAGGAGGGCATATCCCGCCAGCTTAAGAATAA     287

Qy      800    TCCAGAGGGCTGTGAAGCTCTGGGTGAGTGTCGTGGANGCCCAAGTCAGTCGCTACGFTGCTA     859

Db      288    CCATTAAAGAACGTGCGCCAAAACACTCAGCCCGACAACAGAAATCAAATCAGTTTTCGTGGCGGT     347

Qy      860    CTGATTCCCGAGAGTTTATGTGCTCGAGCTCCACACAGCTCTTCAAAGGGAAGG---TGAAGG     916

Db      348    CAGACTCCAAATCACATGATTGGTGAACATAAACACGGCCCCCTTAGTCTGCATGGGCATCACTG     407

Qy      917    TGCTGAGCTCTGAAGCCTGAGGTGGCCAGCGTGCAGCTGTATACATCTCTCGCCCAAGCCGACC     976

Db      408    TGCACAGCTGACGAGGATGATCTTACTCTGAGCTTGGGCCATCTTCGGACAGTCTGAAACC     467

Qy      977    ACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCTCTTGTGAAGCGGGAGCGGACCTCC     1036

Db      468    AC'TTTATCGCACTGATATATCTCTTACTCGGCATT'CGTAAAAAGGGAAACGAGATGTGC     527

Qy      1037 AGSGGAGGGCGGTCTTCTTTCTTCGGCATGCACAGG     1071

Db      528    ACGTTTTCCATCGTACTTCTGGGATCCCCCAAG     562

RESULT 12  
US-09-270-767-28905/c  
; Sequence 28905, Application US/09270767  
; Patent No. 6703491  
GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 28905  
; LENGTH: 998  
; TYPE: DNA  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-28905

; NUMBER OF SEQ ID NOS: 207012

TWO PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 08:00:04 ; Search time 1355.3 Seconds  
(without alignments)  
9238.165 Million cell updates/sec

Title: US-09-774-954-1  
Perfect score: 1514  
Sequence: 1 ATGCCCGCGGCTCTGGCA.....TTTTGAAATCATTC 1514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*  
23: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
24: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*  
26: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
27: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
28: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description        |
|------------|--------|-------------|--------|----|--------------------|
| 1          | 1514   | 100.0       | 1514   | 11 | US-09-774-954-1    |
| 2          | 1512.4 | 99.9        | 11284  | 11 | US-09-774-954-5    |
| 3          | 1510.8 | 99.8        | 5218   | 16 | US-10-301-822-168  |
| 4          | 1510.8 | 99.8        | 5218   | 22 | US-10-956-157-4752 |
| 5          | 1447.4 | 95.6        | 5009   | 11 | US-09-774-954-7    |
|            |        |             |        |    | Sequence 1, Appli  |
|            |        |             |        |    | Sequence 5, Appli  |
|            |        |             |        |    | Sequence 168, App  |
|            |        |             |        |    | Sequence 4752, Ap  |
|            |        |             |        |    | Sequence 7, Appli  |

|    |        |      |        |    |                       |                    |
|----|--------|------|--------|----|-----------------------|--------------------|
| 6  | 1163.4 | 76.8 | 1300   | 11 | US-09-774-954-4       | Sequence 4, Appli  |
| 7  | 1100   | 72.7 | 1100   | 11 | US-09-774-954-16      | Sequence 16, Appl  |
| 8  | 608.6  | 40.2 | 10331  | 10 | US-09-764-891-6032    | Sequence 6032, Ap  |
| 9  | 471.4  | 31.1 | 1400   | 22 | US-10-956-157-9986    | Sequence 9986, Ap  |
| 10 | 471.4  | 31.1 | 1506   | 22 | US-10-956-157-4751    | Sequence 4751, Ap  |
| 11 | 442.8  | 29.2 | 477    | 10 | US-09-918-995-22744   | Sequence 22744, A  |
| 12 | 441    | 29.1 | 479    | 10 | US-09-918-995-19437   | Sequence 19437, A  |
| 13 | 166.4  | 11.0 | 1209   | 26 | US-11-097-143-18161   | Sequence 18161, A  |
| 14 | 150.4  | 9.9  | 3264   | 26 | US-11-097-143-18160   | Sequence 18160, A  |
| 15 | 114.6  | 7.6  | 402    | 9  | US-09-783-590-11501   | Sequence 11501, A  |
| 16 | 108.6  | 7.2  | 3793   | 26 | US-11-097-143-18082   | Sequence 18082, A  |
| 17 | 71.8   | 4.7  | 3567   | 26 | US-11-097-143-18169   | Sequence 18169, A  |
| 18 | 60     | 4.0  | 60     | 10 | US-09-908-975-17287   | Sequence 17287, A  |
| 19 | 56     | 3.7  | 760    | 24 | US-10-820-474A-236    | Sequence 236, App  |
| 20 | 55.6   | 3.7  | 521    | 13 | US-09-925-065A-60244  | Sequence 60244, A  |
| 21 | 42.4   | 2.8  | 1438   | 18 | US-10-148-806-3       | Sequence 19431, A  |
| 22 | 41.2   | 2.7  | 114793 | 16 | US-10-148-806-3       | Sequence 3, Appli  |
| 23 | 41.2   | 2.7  | 114793 | 24 | US-10-859-792-3       | Sequence 3, Appli  |
| 24 | 40.6   | 2.7  | 2036   | 18 | US-10-369-493-43154   | Sequence 43154, A  |
| 25 | 40.4   | 2.7  | 1212   | 21 | US-10-425-115-114097  | Sequence 114097, A |
| 26 | 40.4   | 2.7  | 5932   | 16 | US-10-205-823-312     | Sequence 312, App  |
| 27 | 40.4   | 2.7  | 5932   | 26 | US-11-051-454-312     | Sequence 312, App  |
| 28 | 40.2   | 2.7  | 726    | 20 | US-10-437-963-19394   | Sequence 19394, A  |
| 29 | 40     | 2.6  | 40     | 11 | US-09-774-954-10      | Sequence 10, Appl  |
| 30 | 40     | 2.6  | 40     | 11 | US-09-774-954-12      | Sequence 12, Appl  |
| 31 | 40     | 2.6  | 2032   | 20 | US-10-437-963-8098    | Sequence 8098, Ap  |
| 32 | 39.2   | 2.6  | 618    | 13 | US-09-925-065A-571171 | Sequence 571171, A |
| 33 | 39     | 2.6  | 372    | 18 | US-10-282-122A-31476  | Sequence 31476, A  |
| 34 | 39     | 2.6  | 1209   | 10 | US-09-948-820-37      | Sequence 37, Appl  |
| 35 | 39     | 2.6  | 1209   | 21 | US-10-613-076-37      | Sequence 37, Appl  |
| 36 | 39     | 2.6  | 1555   | 24 | US-10-220-335-120     | Sequence 120, App  |
| 37 | 39     | 2.6  | 1575   | 19 | US-10-072-012-131     | Sequence 131, App  |
| 38 | 39     | 2.6  | 1654   | 24 | US-10-220-335-121     | Sequence 121, App  |
| 39 | 39     | 2.6  | 1963   | 24 | US-10-220-335-464     | Sequence 464, App  |
| 40 | 39     | 2.6  | 1963   | 24 | US-10-220-335-465     | Sequence 465, App  |
| 41 | 38.6   | 2.5  | 407    | 9  | US-09-960-352-11732   | Sequence 11732, A  |
| 42 | 38.6   | 2.5  | 2358   | 10 | US-09-919-039-190     | Sequence 190, App  |
| 43 | 38.6   | 2.5  | 2384   | 22 | US-10-278-698-123     | Sequence 123, App  |
| 44 | 38.6   | 2.5  | 2384   | 22 | US-10-278-698-638     | Sequence 638, App  |
| 45 | 38.6   | 2.5  | 2670   | 17 | US-10-247-671-64      | Sequence 64, Appl  |

ALIGNMENTS

RESULT 1  
US-09-774-954-1  
; Sequence 1, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498

```
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

Query Watch      100.0%; Score 1514; DB 11; Length 1514;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCGGGGCTCTGGACCGGGCGGTACTGCTCTACTGCCCCCTGCATGGGGGC 60
Db 1 ATGCCCGGGGCTCTGGACCGGGCGGTACTGCTCTACTGCCCCCTGCATGGGGGC 60

Qy 61 TTGGGAACACAGCCGATCACTCTTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 120
Db 61 TTGGGAACACAGCCGATCACTCTTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 120

Qy 121 ACCTTGGCTGTCCTCTCTGGATTGAGTACCAGATCACAAAGCTCTCTTTCACCAACCTC 180
Db 121 ACCTTGGCTGTCCTCTCTGGATTGAGTACCAGATCACAAAGCTCTCTTTCACCAACCTC 180

Qy 181 CATGTGTCTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTTACCATCGGTCATC 240
Db 181 CATGTGTCTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTTACCATCGGTCATC 240

Qy 241 AGCTTGGAGGATTTCTATGAGAGAGCTGSCACCCACCTGCGCCCTGAGAAAGCGGTG 300
Db 241 AGCTTGGAGGATTTCTATGAGAGAGCTGSCACCCACCTGCGCCCTGAGAAAGCGGTG 300

Qy 301 GCATCTGCTTTGAGTGGAGCCGACAGCGAGCCAGATGAAGAGAGCTGCGCATGAG 360
Db 301 GCATCTGCTTTGAGTGGAGCCGACAGCGAGCCAGATGAAGAGAGCTGCGCATGAG 360

Qy 361 GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTTCAATGAGTTTCAACAGTCCGAG 420
Db 361 GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTTCAATGAGTTTCAACAGTCCGAG 420

Qy 421 CTTTTCACAGCATTTCTTTCAGTGTCTTCTACAGAAACAATGGAGCCAGAGATTTTCT 480
Db 421 CTTTTCACAGCATTTCTTTCAGTGTCTTCTACAGAAACAATGGAGCCAGAGATTTTCT 480

Qy 481 CCAAGGAACAATCCGGTGTCTTGCCTTGCAGGAGCCCGCCAGTTCCTCCCTCTAGAA 540
Db 481 CCAAGGAACAATCCGGTGTCTTGCCTTGCAGGAGCCCGCCAGTTCCTCCCTCTAGAA 540

Qy 541 GAACACAGGCCACTACAGAACTACATGTTGTTGTCAGACGAAATGTTGAAGACGGAG 600
Db 541 GAACACAGGCCACTACAGAACTACATGTTGTTGTCAGACGAAATGTTGAAGACGGAG 600

Qy 601 GCCAGATTATGCCACCTTCTCGGCCCTTATGTTGGGCTTCTATCTGGCATTTGGCTCT 660
Db 601 GCCAGATTATGCCACCTTCTCGGCCCTTATGTTGGGCTTCTATCTGGCATTTGGCTCT 660

Qy 661 GACTGGAAGACGCTTGTGCATCTGTAAGACGGGACTGCGAGCTCGCACTTCATGGCC 720
Db 661 GACTGGAAGACGCTTGTGCATCTGTAAGACGGGACTGCGAGCTCGCACTTCATGGCC 720

Qy 721 TCTCCGAGTGTGGGCTACAGCGCAGCACAGCGGCCCTCCCTACGATGATATGTC 780
Db 721 TCTCCGAGTGTGGGCTACAGCGCAGCACAGCGGCCCTCCCTACGATGATATGTC 780
```

```
Qy 781 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCC 840
Db 781 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCC 840

Qy 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGCTCTTC 900
Db 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGCTCTTC 900

Qy 901 AAAGGGAAGGTGAAGGTGTGAGCTTGAAGCTCTGAGGTGGCCAGGTGACCTGTATCATC 960
Db 901 AAAGGGAAGGTGAAGGTGTGAGCTTGAAGCTCTGAGGTGGCCAGGTGACCTGTATCATC 960

Qy 961 CTCGGCCAAAGCCGACCACTTTATTGGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db 961 CTCGGCCAAAGCCGACCACTTTATTGGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCT 1020

Qy 1021 CGGAGCGGAGACTCCAGGGGAGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1080
Db 1021 CGGAGCGGAGACTCCAGGGGAGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1080

Qy 1081 CTGCGGAGCGAGTTCTGATTTCTGGCGGAGCACAGACCTCTCTGATCTCTGGAGGACGAG 1140
Db 1081 CTGCGGAGCGAGTTCTGATTTCTGGCGGAGCACAGACCTCTCTGATCTCTGGAGGACGAG 1140

Qy 1141 AGTCTGAGCTGCTCTTTCAGCCAGGCTCTGGAGCGCAGAGGTGCTCCGGGATTCGAAACT 1200
Db 1141 AGTCTGAGCTGCTCTTTCAGCCAGGCTCTGGAGCGCAGAGGTGCTCCGGGATTCGAAACT 1200

Qy 1201 CCTCTTCTACCTGCCAAAGATGGAGAGGTGCGAGGACCCCTCAAGGAGGAGAGCGC 1260
Db 1201 CCTCTTCTACCTGCCAAAGATGGAGAGGTGCGAGGACCCCTCAAGGAGGAGAGCGC 1260

Qy 1261 TCCATATCCAGGCGCATAGGACTTTCAGGTTCTTAGAGCAGGAGCATCTCCCATCGCAC 1320
Db 1261 TCCATATCCAGGCGCATAGGACTTTCAGGTTCTTAGAGCAGGAGCATCTCCCATCGCAC 1320

Qy 1321 GTGCTTCTGCTCTTCTTGGGAATTTCTCACACTGGCAAGCAGTCCAGGCTCGCTCTTCT 1380
Db 1321 GTGCTTCTGCTCTTCTTGGGAATTTCTCACACTGGCAAGCAGTCCAGGCTCGCTCTTCT 1380

Qy 1381 GGTCCACTCTCTCTGAGCAGCTCGGATGCTGACACTTTCAGAGATTTTATATAGA 1440
Db 1381 GGTCCACTCTCTCTGAGCAGCTCGGATGCTGACACTTTCAGAGATTTTATATAGA 1440

Qy 1441 GAGATTTCTATATTTTGTATACAGGTCTATGACTATCTAGAACTCTCTGTGTTTGA 1500
Db 1441 GAGATTTCTATATTTTGTATACAGGTCTATGACTATCTAGAACTCTCTGTGTTTGA 1500

Qy 1501 AAATCATTGAATTC 1514
Db 1501 AAATCATTGAATTC 1514
```

```
RESULT 2
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
```

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/774,954  
FILING DATE: 30-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,741  
FILING DATE: 26-Nov-1997  
APPLICATION NUMBER: 08/792,498  
FILING DATE: 31-Jan-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Svoboda, Craig G.  
REGISTRATION NUMBER: 39,044  
REFERENCE/DOCKET NUMBER: P1041P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1489  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11284 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-774-954-5

Query Match 99.9%; Score 1512.4; DB 11; Length 11284;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

|    |      |   |      |
|----|------|---|------|
| QY | 1    | ATGCGCGGGGCTCTGGGACCGCGCGGTACCTGCTCTACTGCGCCCTGCATGGGGGC  | 60   |
| DB | 4236 | ATGCGCGGGGCTCTGGGACCGCGCGGTACCTGCTCTACTGCGCCCTGCATGGGGGC  | 4295 |
| QY | 61   | TTTGGGAACCGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT | 120  |
| DB | 4296 | TTTGGGAACCGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT | 4355 |
| QY | 121  | ACCTTGGGTCCTCTCTGGATGAGTACAGCATCAAGCCCTCTCTTACCAACCTC     | 180  |
| DB | 4356 | ACCTTGGGTCCTCTCTGGATGAGTACAGCATCAAGCCCTCTCTTACCAACCTC     | 4415 |
| QY | 181  | CATGTGCTTACAGAACTCTTAAGCTGGAGCCCTCCAGGCTTACATGGGTATC      | 240  |
| DB | 4416 | CATGTGCTTACAGAACTCTTAAGCTGGAGCCCTCCAGGCTTACATGGGTATC      | 4475 |
| QY | 241  | AGCTTGGAGGATTTTCATGAGAGCTGGCACCACCTGCGCCCTGAGAGCGGGTG     | 300  |
| DB | 4476 | AGCTTGGAGGATTTTCATGAGAGCTGGCACCACCTGCGCCCTGAGAGCGGGTG     | 4535 |
| QY | 301  | GCATACCTGTTGAGGTGGCAGCCAGGAGCCAGATAGAGACCGTGCCCATGAAG     | 360  |
| DB | 4536 | GCATACCTGTTGAGGTGGCAGCCAGGAGCCAGATAGAGACCGTGCCCATGAAG     | 4595 |
| QY | 361  | GAAGAAACCCCTTGGCCCATCTGGGATCAGATTTTCAATGAGTTTCAAGTCGGAG   | 420  |
| DB | 4596 | GAAGAAACCCCTTGGCCCATCTGGGATCAGATTTTCAATGAGTTTCAAGTCGGAG   | 4655 |
| QY | 421  | CTTTTACAGCATTTCTCTGAGTCTTCTACAGAACCAATGGAGCCAGATTTTCT     | 480  |
| DB | 4656 | CTTTTACAGCATTTCTCTGAGTCTTCTACAGAACCAATGGAGCCAGATTTTCT     | 4715 |
| QY | 481  | CCAAAGAACATCCGGTCTTGGCCCTGCGAGGCCCCAGCCAGTTCCTCTAGAA      | 540  |
| DB | 4716 | CCAAAGAACATCCGGTCTTGGCCCTGCGAGGCCCCAGCCAGTTCCTCTAGAG      | 4775 |
| QY | 541  | GAACACAGGCACTACAGAACTACATGTATGGTTCAGAGAAATGGTGAAGACGGAGAG | 600  |
| DB | 4776 | GAACACAGGCACTACAGAACTACATGTATGGTTCAGAGAAATGGTGAAGACGGAGAG | 4835 |
| QY | 601  | GCCAGATTATGCCACCTTGTCCGGCCCTATGTGGGCAATCTCTGGCAATGGCTCT   | 660  |
| DB | 4836 | GCCAGATTATGCCACCTTGTCCGGCCCTATGTGGGCAATCTCTGGCAATGGCTCT   | 4895 |

|    |      |  |      |
|----|------|--|------|
| QY | 661  | GACTGGAAGAACGCTCTGTGCCATGCTGAAGGACGGGACTGACGGCTCGCACTTATGGGC | 720  |
| DB | 4896 | GACTGGAAGAACGCTCTGTGCCATGCTGAAGGACGGGACTGACGGCTCGCACTTATGGGC | 4955 |
| QY | 721  | TCTCGCAGTGTGTGGGTACAGCGCAGCACAGCGGCCCCCTCACGATGATATGTGC      | 780  |
| DB | 4956 | TCTCGCAGTGTGTGGGTACAGCGCAGCACAGCGGCCCCCTCACGATGATATGTGC      | 5015 |
| QY | 781  | CTGCGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGTGGTGGTGGATGCC     | 840  |
| DB | 5016 | CTGCGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGTGGTGGTGGATGCC     | 5075 |
| QY | 841  | CAGTCGCTCTACGTTGCTACTGATTCGAGAGTATATGTGCTGAGCTCCAAAGCTCTTC   | 900  |
| DB | 5076 | CAGTCGCTCTACGTTGCTACTGATTCGAGAGTATATGTGCTGAGCTCCAAAGCTCTTC   | 5135 |
| QY | 901  | AAAGGGAAGGTGAAGGTGTGAGCTGAGGCTGAGGTGGCCAGGTCGACTGTATCATC     | 960  |
| DB | 5136 | AAAGGGAAGGTGAAGGTGTGAGCTGAGGCTGAGGTGGCCAGGTCGACTGTATCATC     | 5195 |
| QY | 961  | CTGCGCCCAAGCGGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTTGAAG   | 1020 |
| DB | 5196 | CTGCGCCCAAGCGGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTTGAAG   | 5255 |
| QY | 1021 | CGGAGCGGACCTCCAGGGAGGCGCTTCTTTCTTTCGGCATGACAGGCGCCCTTAAG     | 1080 |
| DB | 5256 | CGGAGCGGACCTCCAGGGAGGCGCTTCTTTCTTTCGGCATGACAGGCGCCCTTAAG     | 5315 |
| QY | 1081 | CTGCGGGAAGTGTCTGATTTCTGGCCGAGCACCAAGCCCTCTGATCTCTGGAGGACCA   | 1140 |
| DB | 5316 | CTGCGGGAAGTGTCTGATTTCTGGCCGAGCACCAAGCCCTCTGATCTCTGGAGGACCA   | 5375 |
| QY | 1141 | AGTCTGAGCTGCTCTTCCAGCCAGGCTGCGACGAGGCTCTCCGGGATTCGAACT       | 1200 |
| DB | 5376 | AGTCTGAGCTGCTCTTCCAGCCAGGCTGCGACGAGGCTCTCCGGGATTCGAACT       | 5435 |
| QY | 1201 | CCTCTTCTCACTGCAAGATGGAAGAGAGTGGCAGGAGCCCTTCAAGGAGGAGAGCG     | 1260 |
| DB | 5436 | CCTCTTCTCACTGCAAGATGGAAGAGAGTGGCAGGAGCCCTTCAAGGAGGAGAGCG     | 5495 |
| QY | 1261 | TCCATATCCAGGCAATAGGACTTGTGAGGTTCTTAGGAGGAGGAGCATCTCCCATCGAC  | 1320 |
| DB | 5496 | TCCATATCCAGGCAATAGGACTTGTGAGGTTCTTAGGAGGAGGAGCATCTCCCATCGAC  | 5555 |
| QY | 1321 | GTGCTTCTGCTCTTCTGGGAAATTTCTCAGACTGGCAAGCAGTCCAGCTCCGCTTCT    | 1380 |
| DB | 5556 | GTGCTTCTGCTCTTCTGGGAAATTTCTCAGACTGGCAAGCAGTCCAGCTCCGCTTCT    | 5615 |
| QY | 1381 | GGTCCACTCTGCTGAGCAGCTGGGATGCTGAACTCTTTCAGAGAGATTTTTTATAGA    | 1440 |
| DB | 5616 | GGTCCACTCTGCTGAGCAGCTGGGATGCTGAACTCTTTCAGAGAGATTTTTTATAGA    | 5675 |
| QY | 1441 | GAGATTTCTAATTTTGTATCAAGGTCATGACTATCTAGAACTCTCTGGTGGTTTTGA    | 1500 |
| DB | 5676 | GAGATTTCTAATTTTGTATCAAGGTCATGACTATCTAGAACTCTCTGGTGGTTTTGA    | 5735 |
| QY | 1501 | AAATCATTTGAATTC 1514   |      |
| DB | 5736 | AAATCATTTGAATTC 5749   |      |

RESULT 3  
US-10-821-168  
; Sequence 168, Application US/10301822  
; Publication No. US20030148410A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: Berge, Allison  
; APPLICANT: Guillemette, Tracy L.  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Thibodeau, Stephen N.





ATTORNEY/AGENT INFORMATION :

NAME: Svoboda, Craig G.  
REGISTRATION NUMBER: 39,044  
REFERENCE/DOCKET NUMBER: P1041P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1489  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5009 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-774-954-7

Query Match 95.6%; Score 1447.4; DB 11; Length 5009;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 66 GAACAGGCGGATCACTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGTACCTT 125  
Db |||||  
Qy 1 GAACAGGCGGATCACTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGTACCTT 60  
Db |||||  
Qy 126 GGCTGTCCCTCTTGGATTGAGTACCAGCATCACAAGCCTCTCTTTCACCAACCTCCATGT 185  
Db |||||  
Qy 61 GCGCTGTCCCTCTTGGATTGAGTACCAGCATCACAAGCCTCTCTTTCACCAACCTCCATGT 120  
Db |||||  
Qy 186 GTCTTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCAATCAGCTT 245  
Db |||||  
Qy 121 GTCTTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCAATCAGCTT 180  
Db |||||  
Qy 246 GGAGATTTATGAGAGAGCTGGACCCACCTGCGCCCTGAGAGCGGGTGGCATA 305  
Db |||||  
Qy 181 GGAGATTTATGAGAGAGCTGGACCCACCTGCGCCCTGAGAGCGGGTGGCATA 240  
Db |||||  
Qy 306 CTGCTTTGAGTGGAGCGGAGCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365  
Db |||||  
Qy 241 CTGCTTTGAGTGGAGCGGAGCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Db |||||  
Qy 366 AAACCCCTTTGGCCATCTTGGATCAGTTTCAATGAGTTTCAAGTGGAGCTTTT 425  
Db |||||  
Qy 301 AAACCCCTTTGGCCATCTTGGATCAGTTTCAATGAGTTTCAAGTGGAGCTTTT 360  
Db |||||  
Qy 426 TACAGGCAATTTCTTCAAGTCTTCTTACAGAGAACTTGGAGCGAGAGATTTTCTCCAAA 485  
Db |||||  
Qy 361 TACAGGCAATTTCTTCAAGTCTTCTTACAGAGAACTTGGAGCGAGAGATTTTCTCCAAA 420  
Db |||||  
Qy 486 GGAACATCCGGTGTCTTGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545  
Db |||||  
Qy 421 GGAACATCCGGTGTCTTGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Db |||||  
Qy 546 CAGGCACTTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605  
Db |||||  
Qy 481 CAGGCACTTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
Db |||||  
Qy 606 GATTCACTCCCACTTCTTGGCCCTTATGAGGCAATTCATCTGCGATTGGCTCTGACTG 665  
Db |||||  
Qy 541 GATTCACTCCCACTTCTTGGCCCTTATGAGGCAATTCATCTGCGATTGGCTCTGACTG 600  
Db |||||  
Qy 666 GAAGAACGCTGTGCGATCTGAG 725  
Db |||||  
Qy 601 GAAGAACGCTGTGCGATCTGAG 660  
Db |||||  
Qy 726 GCAGTGTGGGCTACAGGCGAG 785  
Db |||||  
Qy 661 GCAGTGTGGGCTACAGGCGAG 720  
Db |||||  
Qy 786 TGACCTGGAAGGAGATCCAG 845  
Db |||||  
Qy 721 TGACCTGGAAGGAGATCCAG 780  
Db |||||  
Qy 846 GGTCTACGTTGCTACTGATCCAGAGATTTATGTGCTGAGCTCCACAGCTCTTCAAAGG 905  
Db |||||

Db 781 GGTCTACGTTGCTACTGATTCGAGAGATTATGTGCTGAGCTCCAAACGCTCTTCAAGG 840  
Qy 906 GAAGTGAAGTGGTGGCTGAGCTGAAGCTGAGGTGGCCAGGTGAGCTTATCATCTCGG 965  
Db 841 GAAGTGAAGTGGTGGCTGAGCTGAAGCTGAGGTGGCCAGGTGAGCTTATCATCTCGG 900  
Qy 966 CAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1025  
Db 901 CCAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
Qy 1026 GGGGACCTCCAGGGAGAGCGGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1085  
Db 961 GGGGACCTCCAGGGAGAGCGGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1020  
Qy 1086 GGACGAGTTCTGATTTGCGCGAGACACAGACCTCTGATCTCTGAGGAGACAGAGTCT 1145  
Db 1021 GGACGAGTTCTGATTTGCGCGAGACACAGACCTCTGATCTCTGAGGAGACAGAGTCT 1080  
Qy 1146 GAGCTGTCTCTTCCAGCGAGGCTGGCAGCGAGAGTGTCTCCGGATTCGAAACTCTCT 1205  
Db 1081 GAGCTGTCTCTTCCAGCGAGGCTGGCAGCGAGAGTGTCTCCGGATTCGAAACTCTCT 1140  
Qy 1206 TCTCACTGCCAAAGATGAGAGAGTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1265  
Db 1141 TCTCACTGCCAAAGATGAGAGAGTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200  
Qy 1266 ATCCAGGCGCATAGGATTTGAGGTTCTTAGAGAGAGAGATCTCCCATTCGACGCTGT 1325  
Db 1201 ATCCAGGCGCATAGGATTTGAGGTTCTTAGAGAGAGAGATCTCCCATTCGACGCTGT 1260  
Qy 1326 TTCTGCTCTCTTCTGGGAATTTCTCACACTGGCAAGAGTCCAGGCTCCGCTCTCTTGGTCC 1385  
Db 1261 TTCTGCTCTCTTCTGGGAATTTCTCACACTGGCAAGAGTCCAGGCTCCGCTCTCTTGGTCC 1320  
Qy 1386 ACTCTGCTCTCAGCAGGCTGGGATGCTGAACTCTTCTCAGAGAGATTTTATATAGAGAT 1445  
Db 1321 ACTCTGCTCTCAGCAGGCTGGGATGCTGAACTCTTCTCAGAGAGATTTTATATAGAGAT 1380  
Qy 1446 TTCTATATTTTGTATACAGTCACTGATCTCTTCTCAGAGATTTTATATAGAGAT 1505  
Db 1381 TTCTATATTTTGTATACAGTCACTGATCTCTTCTCAGAGATTTTATATAGAGAT 1440  
Qy 1506 ATTGAATTC 1514  
Db 1441 ATTGAATTC 1449

RESULT 6  
US-09-774-954-4  
; Sequence 4, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Winpatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741

|   |                                |                   |
|---|--------------------------------|-------------------|
| : | FILING DATE:                   | 26-NOV-1997       |
| : | APPLICATION NUMBER:            | 08/792,498        |
| : | FILING DATE:                   | 31-JAN-1997       |
| : | ATTORNEY/AGENT INFORMATION:    |                   |
| : | NAME:                          | Svoboda, Craig G. |
| : | REGISTRATION NUMBER:           | 39,044            |
| : | REFERENCE/DOCKET NUMBER:       | P1041PI           |
| : | TELECOMMUNICATION INFORMATION: |                   |
| : | TELEPHONE:                     | 650/225-1489      |
| : | TELEFAX:                       | 650/952-9881      |
| : | INFORMATION FOR SEQ ID NO: 4:  |                   |
| : | SEQUENCE CHARACTERISTICS:      |                   |
| : | LENGTH:                        | 1300 base pairs   |
| : | TYPE:                          | Nucleic Acid      |
| : | STRANDEDNESS:                  | Single            |
| : | TOPOLOGY:                      | Linear            |
| : | SEQUENCE DESCRIPTION:          | SEQ ID NO: 4:     |
| : | US-09-774-954-4                |                   |

  

|                       |                 |               |           |              |
|-----------------------|-----------------|---------------|-----------|--------------|
| Query Match           | 76.8%;          | Score 1163.4; | DB 11;    | Length 1300; |
| Best Local Similarity | 99.9%;          | Pred. No. 0;  |           |              |
| Matches 1164;         | Conservative 0; | Mismatches 1; | Indels 0; | Gaps 0       |

  

|    |     |   |     |
|----|-----|---|-----|
| Qy | 1   | ATGCCCCGGGCTCCTGGAGCCCGGCCGTTCACCTGCTACTGCTTACTGCCCCTGCATGGGGGCGC | 60  |
| Dd | 136 |   |     |
| Qy | 61  | TTTGGGAACCAGGCGCATCACTTCTTGCGCTCTCTGGCATTTGAAAGCTGCTAAACCGT       | 120 |
| Dd | 196 | TTTGGGAACCAGGCGCATCACTTCTTGCGCTCTCTGGCATTTGAAAGCTGCTAAACCGT       | 255 |
| Qy | 121 | ACCTTGCGTGTCCCTCTTGATTTGAGTACCAGCATCAACAAGCCTCTTTTACCAACATCTC     | 180 |
| Dd | 256 | ACCTTGCGTGTCCCTCTTGATTTGAGTACCAGCATCAACAAGCCTCTTTTACCAACATCTC     | 315 |
| Qy | 181 | CATGTGTCCTACCAAGAAGTACTTCAAAGCTGAGGCCCCCTCCAGGCTTACCATCGGTCATC    | 240 |
| Dd | 316 | CATGTGTCCTACCAAGAAGTACTTCAAAGCTGAGGCCCCCTCCAGGCTTACCATCGGTCATC    | 375 |
| Qy | 241 | AGCTTGAGGAGATTTTCATGGAGAAGCTGSCACCCACCCACTGGCCCCCTGAGAAGCGGGTG    | 300 |
| Dd | 376 | AGCTTGAGGAGATTTTCATGGAGAAGCTGSCACCCACCCACTGGCCCCCTGAGAAGCGGGTG    | 435 |
| Qy | 301 | GCATACTGCTTTGAGGTGGCAGCCCGAAGCCCGAGATAAGAGACGTGCCCCCATGAAG        | 360 |
| Dd | 436 | GCATACTGCTTTGAGGTGGCAGCCCGAAGCCCGAGATAAGAGACGTGCCCCCATGAAG        | 495 |
| Qy | 361 | GAAGGAACCCCTTTGGGCCATTTCTGGGATCAGTTTCATGTGAGTTTCAAACAAGTCGGAG     | 420 |
| Dd | 496 | GAAGGAACCCCTTTGGGCCATTTCTGGGATCAGTTTCATGTGAGTTTCAAACAAGTCGGAG     | 555 |
| Qy | 421 | CTTTTTACAGGCATTTCTTTCAGTGCTTTCCTPACAGAGAACAAATGGAGCCACAGATTTTCT   | 480 |
| Dd | 556 | CTTTTTACAGGCATTTCTTTCAGTGCTTTCCTPACAGAGAACAAATGGAGCCACAGATTTTCT   | 615 |
| Qy | 481 | CCAAGGAACATCCGGTGCTTGCCCTGCCAGAGCCCCCAGGCCCTAGTTCCCCGTCTAGAA      | 540 |
| Dd | 616 | CCAAGGAACATCCGGTGCTTGCCCTGCCAGAGCCCCCAGGCCCTAGTTCCCCGTCTAGAG      | 675 |
| Qy | 541 | GAACACAGGGCACTACAGAGATGATCATGGTATGGTTCAGACGAAATGGTGAAGACGGGAGAG   | 600 |
| Dd | 676 | GAACACAGGGCACTACAGAGATGATCATGGTATGGTTCAGACGAAATGGTGAAGACGGGAGAG   | 735 |
| Qy | 601 | GCCCAGATTCATGCCACCTTGTCGGCCCTCATGTGGGCATTCATCTGGCGCATTTGGCTCT     | 660 |
| Dd | 736 | GCCCAGATTCATGCCACCTTGTCGGCCCTCATGTGGGCATTCATCTGGCGCATTTGGCTCT     | 795 |
| Qy | 661 | GACTGAAGAAACGCTCTGTCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGSCC       | 720 |
| Dd | 796 | GACTGAAGAAACGCTCTGTCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGSCC       | 855 |
| Qy | 721 | TCTCCGAGTGTGTGGGCTAAGCCGCAGCACAGCGGCCCCCCCTTCACGATGACTATGTGC      | 780 |
|    |     |   |     |

SEQUENCE DESCRIPTION: SEQ ID NO: 16;

US-09-774-954-16

Query Match 72.7%; Score 1100; DB 11; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |     |   |      |
|----|-----|---|------|
| Qy | 1   | ATGCCCGGGCTCTGGGACCGCGCGTGTACCTGCTTACTGCGCCCTGCGATGGGCGC      | 60   |
| Db | 1   | ATGCCCGGGCTCTGGGACCGCGCGTGTACCTGCTTACTGCGCCCTGCGATGGGCGC      | 60   |
| Qy | 61  | TTTGGGACACGCGCGATCACTTCTTGGGCTCTGCGCATTTGCAAGCTGTAACCGT       | 120  |
| Db | 61  | TTTGGGACACGCGCGATCACTTCTTGGGCTCTGCGCATTTGCAAGCTGTAACCGT       | 120  |
| Qy | 121 | ACCTTGGCTGTCCTCTTGGATGAGTACAGCATCACAAAGCTCTTTACCAACCTC        | 180  |
| Db | 121 | ACCTTGGCTGTCCTCTTGGATGAGTACAGCATCACAAAGCTCTTTACCAACCTC        | 180  |
| Qy | 181 | CATGTGCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGTCATC     | 240  |
| Db | 181 | CATGTGCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGTCATC     | 240  |
| Qy | 241 | AGCTTGGAGGATTTTCATGAGAGTGGCACCCACCTGGCGCCCTGAGAAGCGGGTG       | 300  |
| Db | 241 | AGCTTGGAGGATTTTCATGAGAGTGGCACCCACCTGGCGCCCTGAGAAGCGGGTG       | 300  |
| Qy | 301 | GCATACTGCTTTGAGGTGGCAGCCAGAGGAGCCAGATTAAGAGACGTGCCCCATGAAG    | 360  |
| Db | 301 | GCATACTGCTTTGAGGTGGCAGCCAGAGGAGCCAGATTAAGAGACGTGCCCCATGAAG    | 360  |
| Qy | 361 | GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTAGTTTCAACAGTCCGGAG    | 420  |
| Db | 361 | GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTAGTTTCAACAGTCCGGAG    | 420  |
| Qy | 421 | CTTTTACAGGATTTCTTACGTGCTCTTACAGAGAACATGAGGACGAGATTTTCT        | 480  |
| Db | 421 | CTTTTACAGGATTTCTTACGTGCTCTTACAGAGAACATGAGGACGAGATTTTCT        | 480  |
| Qy | 481 | CCAAAGGAACATCCGGTGCTTGCCTTGCAGAGAGCCCGCCAGTTCCTCCCTCTAGAA     | 540  |
| Db | 481 | CCAAAGGAACATCCGGTGCTTGCCTTGCAGAGAGCCCGCCAGTTCCTCCCTCTAGAA     | 540  |
| Qy | 541 | GAACACAGGCCACTACAGAAATCATGGTATGGTTCAGAGAGGCTGAGGACGGGAG       | 600  |
| Db | 541 | GAACACAGGCCACTACAGAAATCATGGTATGGTTCAGAGAGGCTGAGGACGGGAG       | 600  |
| Qy | 601 | GCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGGCAATGGCTCT | 660  |
| Db | 601 | GCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGGCAATGGCTCT | 660  |
| Qy | 661 | GACTGGAAGAACGCTGTGCTCATGTGAGGACGGGACTGCGAGCTCGCACTTCATGGCC    | 720  |
| Db | 661 | GACTGGAAGAACGCTGTGCTCATGTGAGGACGGGACTGCGAGCTCGCACTTCATGGCC    | 720  |
| Qy | 721 | TCTCCGAGTGTGGGTACAGCGCAGCAGCGGCGCCCTCACGATGACTATGTGC          | 780  |
| Db | 721 | TCTCCGAGTGTGGGTACAGCGCAGCAGCGGCGCCCTCACGATGACTATGTGC          | 780  |
| Qy | 781 | CTGCTGACCTGAAGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTTCGCTGATGCC       | 840  |
| Db | 781 | CTGCTGACCTGAAGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTTCGCTGATGCC       | 840  |
| Qy | 841 | CAGTCGGTCTAGTGTCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC      | 900  |
| Db | 841 | CAGTCGGTCTAGTGTCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC      | 900  |
| Qy | 901 | AAAGGGAAGGTGAAGGTGGTGGCTGAAAGCTGAGGTGGCCAGGTCGACCTGTACATC     | 960  |
| Db | 901 | AAAGGGAAGGTGAAGGTGGTGGCTGAAAGCTGAGGTGGCCAGGTCGACCTGTACATC     | 960  |
| Qy | 961 | CTCGGCCAAGCGGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG    | 1020 |
| Db |     | CTCGGCCAAGCGGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG    |      |

RESULT 8

US-09-764-891-6032  
; Sequence 6032, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6032  
; LENGTH: 10331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6032

|                       |              |   |               |               |         |
|-----------------------|--------------|---|---------------|---------------|---------|
| Query Match           | 40.2%;       | Score 608.6;  | DB 10;        | Length 10331; |         |
| Best Local Similarity | 98.6%;       | Pred. No. 5.4e-186;   |               |               |         |
| Matches 614;          | Conservative | 0;  | Mismatches 9; | Indels 0;     | Gaps 0; |
| QY                    | 892          | CAGCTCTTCAAAAGGGAAGGTGAAGTGGTGAAGCTGAGGTGCCCGAGGTCCAGC        | 951           |               |         |
| Db                    | 6127         | CTGTGCTCTTCTTCGACAGTGAAGTGGTGAAGCTGAGGTGCCCGACAGTCCAGC        | 6186          |               |         |
| QY                    | 952          | CTGTACATCTCTGGGCAAGCGACCACTTTATTGGCAACTGTGTCTCTCTCTCACTGCC    | 1011          |               |         |
| Db                    | 6187         | CTGTACATCTCTGGGCAAGCGACCACTTTATTGGCAACTGTGTCTCTCTCTCACTGCC    | 6246          |               |         |
| QY                    | 1012         | TTTGTGAAGCGGAGCGGGACCTCCAGGGAGGCGCTCTCTTCTTCGCGATGCACAGG      | 1071          |               |         |
| Db                    | 6247         | TTTGTGAAGCGGAGCGGGACCTCCAGGGAGGCGCTCTCTTCTTCGCGATGCACAGG      | 6306          |               |         |
| QY                    | 1072         | CCCCCTAAGCTGCGGGACGAGTTCTGATTTCTGGCGGAGCACAGACCCCTCTGATCTCGG  | 1131          |               |         |
| Db                    | 6307         | CCCCCTAAGCTGCGGGACGAGTTCTGATTTCTGGCGGAGCACAGACCCCTCTGATCTCTGG | 6366          |               |         |
| QY                    | 1132         | AGGGAACACAGAGTCTGAGCTGTGCTTTCACGCCAGGCGCTGGCAGCAGAGGTGCTCGGGA | 1191          |               |         |
| Db                    | 6367         | AGGGAACACAGAGTCTGAGCTGTGCTTTCACGCCAGGCGCTGGCAGCAGAGGTGCTCGGGA | 6426          |               |         |
| QY                    | 1192         | TTGCAAACTCCTCTTCTCACCCTGCAAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA | 1251          |               |         |
| Db                    | 6427         | TTGCAAACTCCTCTTCTCACCCTGCAAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA | 6486          |               |         |
| QY                    | 1252         | GGGAGACGCTCCATATCCAGGGCATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTC   | 1311          |               |         |
| Db                    | 6487         | GGGAGACGCTCCATATCCAGGGCATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTC   | 6546          |               |         |
| QY                    | 1312         | CCATCGACGCTGCTTCTGCTCTTCTGGGAAATTTCTCACTGCTGGCAAGCAGTCCAGCT   | 1371          |               |         |
| Db                    | 6547         | CCATCGACGCTGCTTCTGCTCTTCTGGGAAATTTCTCACTGCTGGCAAGCAGTCCAGCT   | 6606          |               |         |
| QY                    | 1372         | CCGCTTCTGTGTCACCTCTGCTCTGAGCAGCCTGGGATGCTGAATCTCTTCAGAGAGATT  | 1431          |               |         |
| Db                    | 6607         | CCGCTTCTGTGTCACCTCTGCTCTGAGCAGCCTGGGATGCTGAATCTCTTCAGAGAGATT  | 6666          |               |         |
| QY                    | 1432         | TTTTATAGAGAGATTCTTATAATTTTGATACAAAGGTCAATGATCTCTAGAACTCTCTGT  | 1491          |               |         |
| Db                    | 6667         | TTTTATAGAGAGATTCTTATAATTTTGATACAAAGGTCAATGATCTCTAGAACTCTCTGT  | 6726          |               |         |





```
QY 197 ACTACTTCAAGCTGAGGCCCCCAGGCTTACCATCGGTCTATCAGCTTGGAGATTCA 256
  |||
Db 254 CATATTTGAAGTGGAGCCCTGAAGGATACCATCGCTCATCACCATGGCAGATTCA 313
QY 257 TGGAGAGCTGGACCCACCCACTGCGCCCTTGAAGAGCGGTGGATCTGCTTTGAGG 316
  |||
Db 314 TGTGGCACCTGGCCGACGACATTTGGCCAGAAATCGAGCGAGTGTCAITTTGCTACAAGG 373
QY 317 -----TGGCAGCCCGAGGAGCCAGATAAGAGACGCTGCCCATGAAGG 361
  |||
Db 374 AACGATATAGCCTTACGAGGAGAGAACGATCCAGAACGCCCATTGGCCAGCCAAAG 433
QY 362 AAGGAAACCCCTTTGGCCCAITCTGGGATCAGTTTCATGTGAGTTTCAACAGTCCGAGC 421
  |||
Db 434 ATGGCAATCTTTTGGTCCCTTTGGACACTTTTACATAGACTTTTGTGCGGTGAGAT 493
QY 422 TTTTACAGGCAATTCCTTCACTGCTTCTTACAG-----AGAACATGAGCCAGA 472
  |||
Db 494 TCTATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACA 553
QY 473 GATTTTCTCCAAAGGAAATCCGCTGCTTGCCTGCGAGGAGCCCGCCAGTTCCCGG 532
  |||
Db 554 AATATCTTCGAGAAATCATATCCGCTACTGCGGTTACCGGAGCTCCGGCTAGTTTCTCTG 613
QY 533 TCCTAGAGAAACACAGGCCACTACAGAACTACATGTGTGTCAGACGAAATGTTGAAGA 592
  |||
Db 614 TTCAGCTAGAGAACTGCAAGCTGCAAGCTACTTTCAGTGGAGTCAACGATATAGGGAAG 673
QY 593 CGGAGAGGCCAGATTCATGCCCCACCTTGTCCCGC---CTPATGTGGCATTCATCTGC 649
  |||
Db 674 CATCTAAGGATTTTCATCCGAGACGAGTTGCTCGGGGTGCCTTTTGGGCATTCATCTGC 733
QY 650 GCAATGGCTCTGACTGGAAGACGCTGTGCCATGCTGAAGACCG-----GACTG 700
  |||
Db 734 GCAACGGTATCGATTTGGGTGAGAGCTGTGAGACGTCGAAGGATAGCCAGCATCTGTTTG 793
QY 701 CAGGCTCGCACTTCATGGGCTCTCCGCACTGTGGGCTACAGCGCGCAGCAGCGGCC 760
  |||
Db 794 CTTCCGCGCAGTGTGGGCTATAAAATGAACGTGGTGCATCTTACCGGAGCTCTGCA 853
QY 761 CCCTCAGATGATATGCTGCTGCTGCTGCTGAAGGAGATCCAGAGGCTGTGAAGCTCT 820
  |||
Db 854 TGCCCTCCAAAGGAGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTCGCCAAA 913
QY 821 GGGTGAAGTCTGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 880
  |||
Db 914 CTCAGCGGACAAACGAAATCAAAATCAGTTTTCGTGGGCTGAGACTTCCAATCACATGATG 973
QY 881 CTGAGCTCCAAAGCTCTTTCAAAGGAGGTGAGGTGCTGAGCTGAGGCTGAGGTGG 940
  |||
Db 974 GTGAATTAACACAGGCTTGTAGTGCATGGGCTCAGTGTGCAAGCTGCGGAGGATG 1033
QY 941 ---CCAGGTGAGCTGATCTCTGCGCCAAAGCGACCACTTTATTTGGCAACTGTGTCT 997
  |||
Db 1034 ATCTTTACCTGGACTTGGCCATTTCTCGGACAGTTCGAACCACTTTATCGGCAACTGTATAT 1093
QY 998 CTTCTTCACTGCTTTGTGAAGCGGGAGCGGAGCTCCAGGGAGGCGCTTCTTCTTCT 1057
  |||
Db 1094 CTTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGACGGTTTTCATCGTACTTCT 1153
QY 1058 TCGGCATGGACAGG 1071
  |||
Db 1154 GGGGATTTCCCAAG 1167
  |||
```

RESULT 14  
US-11-097-143-18160/c  
; Sequence 18160, Application US/11097143  
; Publication No. US2005020858A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.

```
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18160  
; LENGTH: 3264  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-18160
```

```
Query Match          9.9%; Score 150.4; DB 26; Length 3264;  
Best Local Similarity 50.9%; Pred No. 3.6e-37;  
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;  
  
QY 49 TGCATGGGGCGCTTTGGGAACGAGCCGATCACTTTTGGGCTCTCTGGCAATTTCGAAG 108  
  |||  
Db 2098 TTCAAAGAGCGCTTTGGCAACGAGCCGACCTTCTGGGATCATTTGGCCTTCGCCAAG 2039  
QY 109 CTGCTAAACCGTACTTGGCTGTCCCTCTTGGATTGATGATCAGCATCAGAGCTCCT 168  
  |||  
Db 2038 CGCTTAAATGCACCCCTGATCCTCGCGCTGGGTGGAGT-----ATCGTAGGGGTGAA 1985  
QY 169 TTCACCAACCTCCATGTGTCTTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTAC 228  
  |||  
Db 1984 CTGCGATCCCTCAGTACCGTTCAAACATATTTTGAAGTGGAGCCCTGAGGAATAC 1925  
QY 229 CATCGGTCATCAGCTGGAGATTTTCATGAGAGCTGGCAACCCACCCCTGGCCCCCT 288  
  |||  
Db 1924 CATCGGTCATCAGCTGGAGATTTTCATGAGAGCTGGCAACCCACCCCTGGCCCCAAG 1865  
QY 289 GAGAGCGGGTGGCATACTGCTTTGAG-----TGGCAGCCCGCAGCAAGC 333  
  |||  
Db 1864 TCGAGGCGAGTGTCAITTTGTCTAAGGAACGATATAGCTTTCAGAGGAAGAACGAT 1805  
QY 334 CCAGATAAGAGACGCTGCCCATGAAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAG 393  
  |||  
Db 1804 CCAGACAGCCCATTTGCCAGCCAGGATGCGCATCTTTTGGTCCCTTTGGGACACT 1745  
QY 394 TTTTCATGTAGTTTCAACAGTCCGAGCTTTTTCAGGCAATTTCTTCAGTGTCTTCCTAC 453  
  |||  
Db 1744 TTTTCATGTAGTTTGTGCGGTTCAGAGTTTCATGCGCCACTTCATTTTGTGTGTCATCAT 1685  
QY 454 AG-----AGAACATGAGCGCAGAGATTTTCTCCAAAGGAACATCCCGTGTCTGCC 504  
  |||  
Db 1684 AGCAACGAGGCTGCCAAGTGGCAGACCAAAATATCTTCAGAAATCATATCCCGTACTCGCG 1625  
QY 505 CTGCGAGGAGCCCGAGCCCGCTTCCCGTCTCTAGAGAAACACAGGCCACTACAGAAGTAC 564  
  |||  
Db 1624 TTTACCGGAGCTCCGGCTAGTTTTCCTGTTTCAGCTAGAGAACTGCAAGCTCGAGCGCTAC 1565  
QY 565 ATGGTATGGTCAGACGAAATGGTGAAGACGGGAGGCGCCAGATTTTCATGCCACCTTCTGC 624  
  |||  
Db 1564 TTGCAGTGGAGTCAACGCTATAGGGAAGCATCTAAGGATTTTCATCCGAGAGCAGTTGCT 1505
```

|    |      |   |      |
|----|------|---|------|
| Qy | 625  | CGG---CCCTATGTGGGCATTCACTCTGGGATTTGGCTCTGACTGGGAAGAACCTCTGTGCC  | 681  |
| Db | 1504 | CGGGGTGCCCTTTTGGGCATTTCATCTGCGCAACGGTATTCGATTGGGTGAGAGCCTGTGGAG | 1445 |
| Qy | 682  | ATGCTGAAGGACGG-----GACTGCAAGGCTCGCACTTTCATGAGCTCTCTCCGCAGTGT    | 732  |
| Db | 1444 | CACGCTCAAGSATAGCAGCATCTGTTTGGCTCGCCGCAAGTCTTGGGCTATATAAATGAA    | 1385 |
| Qy | 733  | GTGGGCTACAGCCGACGACAGCGGCCCCCTCAGCATGACTATGTGCTGCTGCCTCACTG     | 792  |
| Db | 1384 | CGTGTGCACTTCTACCCGGAGCTCTGCATGCCCTCCAAAGGAGGCAATCATCCGCCAGCTA   | 1325 |
| Qy | 793  | AAGGAGATCCAGAGGCTGTGAAAGCTCTGGGTGAGTGCCTGGATGCCACAGTCGGTCTTAC   | 852  |
| Db | 1324 | AAGAGACCAATTAGAAAGTGGCGCAAACTCAGCCGGCAACAGAAATCAAAATCAGTTTTC    | 1265 |
| Qy | 853  | GTTCTACTATGATTCGGAGAGTTATGTGCGCTGAGCTCCACAGCTCTTCAAGGGGAAGTGT   | 912  |
| Db | 1264 | GTGGGTGACATCCANATCATGATTTGGTGAATTAACACAGGCCCTTAGTCGATGGGC       | 1205 |
| Qy | 913  | AAGTGTGTAGCTGAGCCTGAGGTGG---CCGAGTTCGACCTGTACATCTCTGGCCAA       | 969  |
| Db | 1204 | ATCAGTGTGCACAAGTCTCCGGAGGATGATCTTACCTGGACCTTGGCCATTCTCGACAG     | 1145 |
| Qy | 970  | GCCACCACTTTATTGGCAACTGTGTCCTCTCTTCACTGCTTTTGTGTAAGCGGAGCGG      | 1029 |
| Db | 1144 | TCGAACCACTTTATCGGCAACTGTATATCTCTTACTCTGGCATTCGAAAAAAGGGAAACGA   | 1085 |
| Qy | 1030 | GACCTCCAGGGGAGGCCGTCTTCTTCTTCGGCATGGACAGG                       | 1071 |
| Db | 1084 | GATGTGCACGGTTTTTCATCTGACTCTCTGGGGATTCGCCAAG                     | 1043 |

RESULT 15  
US-09-783-590-11501  
; Sequence 11501, Application US/09783590  
; Patent No. US20020110950A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2



| Result No. | Query No. | Score  | Match | Query $\phi$ | Length | DB       | ID          | Description |  |
|------------|-----------|--------|-------|--------------|--------|----------|-------------|-------------|--|
|            |           |        |       |              |        |          |             |             |  |
| C          | 1         | 1443.8 | 95.4  | 1555         | 3      | CR617950 | full-length |             |  |
|            | 2         | 892.8  | 59.0  | 1141         | 5      | BX353274 | BX353274    |             |  |
|            | 3         | 892    | 58.9  | 3611         | 3      | AK081059 | Mus muscu   |             |  |
|            | 4         | 888    | 58.7  | 2616         | 3      | AK044629 | Mus muscu   |             |  |
|            | 5         | 859.6  | 56.8  | 889          | 5      | BX345988 | BX345988    |             |  |
|            | 6         | 852.4  | 56.3  | 914          | 5      | BQ677046 | BQ677046    |             |  |
|            | 7         | 743.6  | 49.1  | 1091         | 5      | EM909055 | EM909055    |             |  |
|            | 8         | 718.4  | 47.5  | 841          | 5      | BX345989 | BX345989    |             |  |
| C          | 9         | 717.4  | 47.4  | 719          | 1      | AL135434 | AL135434    |             |  |
|            | 10        | 712.2  | 47.0  | 889          | 4      | B1524031 | B1524031    |             |  |
|            | 11        | 706.6  | 46.7  | 898          | 5      | BX449172 | BX449172    |             |  |
|            | 12        | 683.8  | 45.2  | 704          | 4      | BM475212 | BM475212    |             |  |
|            | 13        | 673.2  | 44.5  | 942          | 7      | CG774833 | CG774833    |             |  |
|            | 14        | 662.2  | 43.7  | 808          | 4      | BQ747760 | BQ747760    |             |  |
|            | 15        | 660.2  | 43.6  | 801          | 5      | BP144264 | BP144264    |             |  |
|            | 16        | 654.4  | 43.2  | 689          | 5      | BM972332 | BM972332    |             |  |
| C          | 17        | 639    | 42.2  | 988          | 5      | BQ964179 | BQ964179    |             |  |
|            | 18        | 636.4  | 42.0  | 874          | 4      | BI757540 | BI757540    |             |  |
|            | 19        | 630.8  | 41.7  | 947          | 4      | B1410783 | B1410783    |             |  |
|            | 20        | 630.4  | 41.6  | 640          | 2      | AW957639 | AW957639    |             |  |
|            | 21        | 625.6  | 41.3  | 809          | 4      | BG473023 | BG473023    |             |  |
|            | 22        | 624    | 41.2  | 915          | 5      | BQ936634 | BQ936634    |             |  |
|            | 23        | 609.8  | 40.3  | 705          | 7      | CN348925 | CN348925    |             |  |
|            | 24        | 607    | 40.1  | 608          | 4      | BM711905 | BM711905    |             |  |





Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details.

```

FEATURES
  source
    1. .3611
      /location/Qualifiers
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="C57BL/6J"
        /db_xref="FANTOM_DB:B930076G19"
        /db_xref="taxon:10090"
        /clone="B930076G19"
        /tissue_type="cerebellum"
        /clone_lib="RIKEN full-length enriched mouse cDNA library"
        /dev_stage="10 days neonate"
      11. .1194
        /notes="protein O-fucosyltransferase 1 (MGD|GI:2153207,
          GB|NM_080463, evidence: BLASTN, 99%, match=1290)
          putative"
  misc_feature

```

| ORIGIN | Query Match           | 58.9%;  | Score 892;          | DB 3;           | Length 3611;      |
|--------|-----------------------|---|---------------------|-----------------|-------------------|
|        | Best Local Similarity | 83.3%;  | Pred. No. 6.3e-236; |                 |                   |
|        | Matches 1551;         | Conservative  | 0;                  | Mismatches 180; | Indels 31; Gaps 2 |
| QY     | 5                     | CGCGGGCTCTCGGACCGCGCGGGTTACCTGCTCTACTGCCCCTGATGGCGGCGCTTGG        | 64                  |                 |                   |
| DB     | 101                   | CAGCGGCTCTGGAGCTGCGCGGTTACCTGCTCTACTGCTCCCTGCATGGGGCGCTTGG        | 160                 |                 |                   |
| QY     | 65                    | GGACACGAGCGCATCACTTTCTGGGCTCTCTGCATTTTCGAAGCTGCTAAACCGTACCT       | 124                 |                 |                   |
| DB     | 161                   | GGACACGAGCTGATCACTTTTGGGCTCCCTGGCATTTTGGGAGCTGCTGAAACGCACT        | 220                 |                 |                   |
| QY     | 125                   | TGGCTGTCCCTCTCTGGATTGATGACAGCATCACAGGCTCTCTTTACCAAACTCCATG        | 184                 |                 |                   |
| DB     | 221                   | TGGCTGTACTCTCATGGATTGAATACAACTACAGGCTCTCTTTACCAAACTCCATG          | 280                 |                 |                   |
| QY     | 185                   | TGTCCTACAGAGTACTTTCAAGCTTGGAGCCCTTCAGGCTTACCATCGGTCATCAGCT        | 244                 |                 |                   |
| DB     | 281                   | TGTCCTACAAAGTACTTTCAAACCTGGAGGCTCTCCAGGCTACCTTCGGGTTCTCAGC        | 340                 |                 |                   |
| QY     | 245                   | TGGAGGATTTTCATGGAGAGCTGGCACCCACCTGCTGGGCGCTCGAGAGCGGGTGGCAT       | 304                 |                 |                   |
| DB     | 341                   | TGGAGGACTTCATGGAAATCTGGCACCTCCCACTGGCCCCCTCGAAGCGAGTGGCAT         | 400                 |                 |                   |
| QY     | 305                   | ACTGCTTTGAGTGGCAGCCCGACGAGCCACGATAGAGAGCTGCCCCATGAGGAGAG          | 364                 |                 |                   |
| DB     | 401                   | ACTGCTTTGAGTGGCAGCCCGACGAGTCTGATAGAGAGCATGTCCCATGAGGAGAG          | 460                 |                 |                   |
| QY     | 365                   | GAACCCCTTTGGCCCATTTCTGGATTCAGTTTCATGATGATTTCAACAGTTCGGAGCTTT      | 424                 |                 |                   |
| DB     | 461                   | GAATTCCTTTTGGGCCATTTCTGGACCAAGTTTCATGATGATTTCAATAGTTCAGAACTGT     | 520                 |                 |                   |
| QY     | 425                   | TTACAGGCGATTTCTTCAGTGTCTCTACAGAGAAACAATGGAGCCAGAGATTTTCTCCAA      | 484                 |                 |                   |
| DB     | 521                   | TCACAGGCGATTTCTTCAGCGCTCTCTACAAAGAAACAATGGAGCCACAGAGATTTCTCCAA    | 580                 |                 |                   |
| QY     | 485                   | AGGAACATCCGGTGTCTTGGCCCTGCCAGGAGCCCCAGCCCATTTCCCGTCTCAAGAAGAC     | 544                 |                 |                   |
| DB     | 581                   | AAGAGCATCTCTGTCTCGCATCTGCTGGGGCCCCAGACAGTGTCTCTCTGGAGGAGC         | 640                 |                 |                   |
| QY     | 545                   | ACAGGCCACTACAGAAGTACATGTTGGTCAGACGAAATGGTGAAGACGGGAGAGGCC         | 604                 |                 |                   |
| DB     | 641                   | ACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAAGACGGGAGAGGCC       | 700                 |                 |                   |
| QY     | 605                   | AGATTATCGCCACCTTGTCTCGGCCCTATGTGGGCATTTCATCTGGGCATTTGGCTCTGACT    | 664                 |                 |                   |
| DB     | 701                   | TGATCAGTGGCCACCTCTGTCTCGGCCCTATGTGGGCATTTTCATCTGGGCATTTGGCTCTGACT | 760                 |                 |                   |
| QY     | 665                   | GGAAGAACGCCCTGTGGCCATCTCTGAAGGACGGGACTGACGGCTCGCACTTCATGAGCCTCTC  | 724                 |                 |                   |

Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE

20530913

11076861

4

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6

(bases 1 to 2616)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Retina RNA was provided by Dr. Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged. Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

FEATURES  
source

Location/Qualifiers  
1. .2616  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:A930028F21"  
/db\_xref="taxon:10090"  
/clone="A930028F21"  
/tissue\_type="retina"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
34. .1215

CDS

/note="unnamed protein product; protein O-fucosyltransferase 1 (MGD|GI:2153207, GB|NM\_080463, evidence: BLASTN, 99%, match=1290) putative"

/codon\_start=1

/protein\_id="BAC32009.1"

/db\_xref="GI:26336653"

/translation="WGAAAWAPPHLLLRASFLILLPLLRGSAGSMDLACGLVLCPC  
CMGRFGNADHFLASLAFKLLNRLTAVPMVEYOHKPPFTNLHVSQYKPKLPLQ  
AYHRVSLSDPMENLAPSHWPKRKYAVCFEVAQORSDDKTCPKMKGNGFPFPDQF  
HVSFNKSLFTGISFSASYKEQWTFKPKHFPVLPALGPAQFPVLHRELRQITMV  
WSDMVRTEALISHLVRYVGVHLIRIGSDWKNAKMLKDGDTAGSHMFSPQCVGYS  
RSTATPLMTMCLPDLKEIORAVTLWRLNARSVYIATDSYVSEIQOLFQDKVRV  
VSLKPEVAQIDLYILGQADHFIGNCVSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE  
F"

ORIGIN

Query Match 58.7%; Score 888; DB 3; Length 2616;  
Best Local Similarity 87.1%; Pred. No. 7.4e-235;  
Matches 975; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
QY 3 GCCCGCGCGCTCTGGGACCCGCGGTTACTGCTCTACTGCCCTGCATGGGCGCTT 62  
DB 120 GTACAGCGGCTCTGGGACCTGCGGTTACTGCTCTACTGCTCTGATGGGCGCTT 179  
QY 63 TGGGAACAGGCGCATCTCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTAC 122  
DB 180 TGGGAACAGGCGCATCTCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGC 239  
QY 123 CTGGCTCTCTCTCTGGATGAGTACAGCATCACAAAGCTCTCTTTCACCAACTCCA 182  
DB 240 CTGGCTCTCTCTCTGGATGAGTACAGCATCACAAAGCTCTCTTTCACCAACTCCA 299  
QY 183 TGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCAGTCCGGTCTAC 242  
DB 300 TGTGCTCTACCAAGTACTTCAAACTGGAGCTCTTCAAGCTTACCAGTCCGGTCTAC 359  
QY 243 CTGGAGGATTTCTGGAGAGAGCTGGCAACCAACCTCTGGGCTCTGAGAACGGGTGGC 302  
DB 360 CTGGAGGATTTCTGGAGAGAGCTGGCAACCAACCTCTGGGCTCTGAGAACGGGTGGC 419  
QY 303 ATACTGCTTGGAGTGGCAGCGCAAGCCAGATAGAGAGAGCTGCCCATGAAGA 362  
DB 420 ATACTGCTTGGAGTGGCAGCGCAAGCTCTGATAGAGAGAGCTGCCCATGAAGA 479  
QY 363 AGGAACCCCTTGGCCATTTCTGGATCAGTTCATGATGATGATGATGATGATGATGAT 422  
DB 480 AGGAACCCCTTGGCCATTTCTGGATCAGTTCATGATGATGATGATGATGATGATGAT 539  
QY 423 TTTTACAGGCTTTCTTCTGCTGCTTCTTCTGCTGCTTCTTCTGCTGCTTCTTCTGCT 482  
DB 540 GTTACAGGCTTTCTTCTGCTGCTTCTTCTGCTGCTTCTTCTGCTGCTTCTTCTGCT 599  
QY 483 AAAGAAACATCGGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542  
DB 600 AAAGAAACATCGGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659  
QY 543 ACAGAGCCACTACAGAGTACATGATGCTCAGACCAATGGTGAAGACGGGAGAGGC 602  
DB 660 ACAGAGGAGCTCAGAAAGTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719  
QY 603 CCAGATTCATGCCCACTTGTCTCGGCTTGTCTGCTGCTTGTCTGCTGCTTGTCTGCT 662  
DB 720 CTTGATCAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779  
QY 663 CTGGAAGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722  
DB 780 CTGGAAGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839  
QY 723 TCCGCGTGTCTGGGCTACAGCCGACAGCGGCTTGTGCTGCTGCTGCTGCTGCTGCT 782  
DB 840 CCCTCAGTGTGTGGGCTATAGCCGACAGCGGCTTGTGCTGCTGCTGCTGCTGCTGCT 899  
QY 783 GCCTGACCTGAAGAGATCCAGAGGCTTGTGAAGCTCTGGGTGAGGTGCTGCTGCTGCT 842  
DB 900 CCCTGACCTGAAGAGATCCAGAGGCTTGTGAAGCTCTGGGTGAGGTGCTGCTGCTGCT 959  
QY 843 GTCGGTCTACTGCTTACTGATTCGAGAGTTATGTGCTGCTGCTGCTGCTGCTGCTGCT 902  
DB 960 ATCGGTCTTACATCGCCACAGACTCTGAGAGCTACGTGTGAGAGATCCAGAGCTCTTCAA 1019





Qy 238 ATCAGCTTGGAGGATTTTCATGGAGAGCTGGGACCCACCCACTGGGCCCTCGAAGCGG 297  
Db |||||  
Qy 192 ATCAGCTTGGAGGATTTTCATGGAGAGCTGGGACCCACCCACTGGGCCCTCGAAGCGG 251  
Db |||||  
Qy 298 GTGGCATATGCTTTTGGAGTGGGAGCCAGGAGCCAGATGAAGAGACGTGCCCCATG 357  
Db |||||  
Qy 252 GTGGCATATGCTTTTGGAGTGGGAGCCAGGAGCCAGATGAAGAGACGTGCCCCATG 311  
Db |||||  
Qy 358 AAGGAAGAAACCCCTTTGGGCCATTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 417  
Db |||||  
Qy 312 AAGGAAGAAACCCCTTTGGGCCATTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 371  
Db |||||  
Qy 418 GAGCTTTTACAGGATTTCTTCAGTGTCTTCTACAGAGAACTATGGAGCCAGAGATTT 477  
Db |||||  
Qy 372 GAGCTTTTACAGGATTTCTTCAGTGTCTTCTACAGAGAACTATGGAGCCAGAGATTT 431  
Db |||||  
Qy 478 TCTCCAAAGGAACATCCGCGTGTGCTGCGAGAGCCCGAGCCAGTTCCCGTCTCTA 537  
Db |||||  
Qy 432 TCTCCAAAGGAACATCCGCGTGTGCTGCGAGAGCCCGAGCCAGTTCCCGTCTCTA 491  
Db |||||  
Qy 538 GAAGAACACAGGCCACTACAGAAATACATGTATGTGTGAGTGTGAGAGCGGGA 597  
Db |||||  
Qy 492 GAGGAACACAGGCCACTACAGAAATACATGTATGTGTGAGTGTGAGAGCGGGA 551  
Db |||||  
Qy 598 GAGGCCAGATTCATGCGCCACTTGTGCGGCCCTATGTGCGCATTCATCTGCGCATTCG 657  
Db |||||  
Qy 552 GAGGCCAGATTCATGCGCCACTTGTGCGGCCCTATGTGCGCATTCATCTGCGCATTCG 611  
Db |||||  
Qy 658 TCTGACTGGAAGACGCTGTGCGCATTCGCTGAAGAGCGGACTGAGGCTCGCACTTCATG 717  
Db |||||  
Qy 612 TCTGACTGGAAGACGCTGTGCGCATTCGCTGAAGAGCGGACTGAGGCTCGCACTTCATG 671  
Db |||||  
Qy 718 GCCTCTCCGCAAGTGTGGGCTACAGCGCAGCAGCGGCCCTCCACGATGACTATG 777  
Db |||||  
Qy 672 GCCTCTCCGCAAGTGTGGGCTACAGCGCAGCAGCGGCCCTCCACGATGACTATG 731  
Db |||||  
Qy 778 TGCCCTGCTGACCTGAAGAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGAT 837  
Db |||||  
Qy 732 TGCCCTGCTGACCTGAAGAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGAT 791  
Db |||||  
Qy 838 GCCAGTGGCTTACGTTGCTACTGATTCGAGAGTATGCTGCTGAGTCTCAACAGCTC 897  
Db |||||  
Qy 792 GCCAGTGGCTTACGTTGCTACTGATTCGAGAGTATGCTGCTGAGTCTCAACAGCTC 851  
Db |||||  
Qy 898 TTCAAA--GGGAGGTGAAGTGGTG--AGCCTGAAGCTGAGTGGGCGCCAGGTGCACT 953  
Db |||||  
Qy 852 TTAAAGGGAGGAGGAGGAGGTTGGTGAACCTGTAAGACCTGTAGGGGGGCCAGGTCCA 911  
Db |||||  
Qy 954 GTACATCTCGGCCCAAGCCGACACTTT--ATTGCAACTGTGTCTCTCTCTCACTG--- 1009  
Db |||||  
Qy 912 GTAAATCTTTCGGAAGCGAACCCCTTTAAATTGGGAACCTGTGGGCCCTCTTCCGGGC 971  
Db |||||  
Qy 1010 CCTTTGTCAAGCGGAGCGGACCTCCAGGGGAGCGCTCTTTT 1055  
Db |||||  
Qy 972 CCTTTGGAAGGAGAACGGACCTCTCAGGGAGGGCGGCTTT 1017  
Db |||||

RESULT 8  
EX345989 841 bp mRNA linear EST 23-APR-2004  
LOCUS  
DEFINITION  
EX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
ACCESSION  
EX345989 Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.  
VERSION  
EX345989.2 GI:46549776  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 841)  
Li W B., Gruber C., Jesse J. and Polayes D.  
Full-length cDNA libraries and normalization

JOURNAL  
COMMENT

Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30377055.  
Genoscope - Centre National de Sequencage  
Contact: Genoscope  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
9980.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06QPl&c=9980.f.

FEATURES  
source

1..841  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DJ008YF12"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/cell\_line="JURKAT"  
/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT  
10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 47.5%; Score 718.4; DB 5; Length 841;  
Best Local Similarity 99.5%; Pred. No. 6.9e-188;  
Matches 730; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  
Qy 1 ATGCCCGGGCTCCTGGGACCCGGCGGTACCTGCTACTGCCCCCTGCATGGGCGC 60  
Db |||||  
Qy 109 ATGCTGCGGGCTCTGGGACCCGGCGGTACCTGCTACTGCCCCCTGCATGGGCGC 168  
Db |||||  
Qy 61 TTGGGAACAGGCGGATCACTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 120  
Db |||||  
Qy 169 TTGGGAACAGGCGGATCACTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 228  
Db |||||  
Qy 121 ACCTGGCTGCTCCCTTGGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 180  
Db |||||  
Qy 229 ACCTGGCTGCTCCCTTGGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 288  
Db |||||  
Qy 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATGGGTCATC 240  
Db |||||  
Qy 289 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATGGGTCATC 348  
Db |||||  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCCACTGGCCCCCTGAGAGCGGGTG 300  
Db |||||  
Qy 349 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCCACTGGCCCCCTGAGAGCGGGTG 408  
Db |||||  
Qy 301 GCATAGCTTTTGGTGGGAGCCAGGAGCCAGGAGCCAGATTAAGAGAGCTGCCCATGAAG 360  
Db |||||  
Qy 409 GCATAGCTTTTGGTGGGAGCCAGGAGCCAGGAGCCAGGAGCCAGATTAAGAGAGCTGCCCATGAAG 467  
Db |||||  
Qy 361 GAAGGAACCCCTTGGGCCATTTCTGGGATCAGTTTCATGAGTTTCAACAGTGGAG 420  
Db |||||  
Qy 468 GAAGGAACCCCTTGGGCCATTTCTGGGATCAGTTTCATGAGTTTCAACAGTGGAG 527  
Db |||||  
Qy 421 CTTTTTACAGGCAATTCCTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 480  
Db |||||  
Qy 528 CTTTTTACAGGCAATTCCTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 587  
Db |||||  
Qy 481 CCAAGGAACATCCGCTGCTTGGCCCTGCCAGGAGCCAGCCAGCTTCCCGTCTTAGAA 540  
Db |||||  
Qy 588 CCAAGGAACATCCGCTGCTTGGCCCTGCCAGGAGCCAGCCAGCTTCCCGTCTTAGAG 647  
Db |||||  
Qy 541 GRACAGGCCACTACAGAGTACATGTTATGTCAGAGCAATGTTGAGAGCGGGAGAG 600  
Db |||||  
Qy 648 GAACAGGCCACTACAGAGTACATGTTATGTCAGAGCAATGTTGAGAGCGGGAGAG 707  
Db |||||



```

QY 601 GCCCAGATTATGCCCACCTTTGTCCGGCCCTATGTGGCCATTCTCTGCGCATTTGGCTCT 660
    |||
Db 708 GSCCAGATTATGCCCACCTTTGTCCGGCCCTATGTGGCCATTCTCTGCGCATTTGGCTCT 767
    |||
QY 661 GACTGGAAGAAGCGCTGTGCCATGCTGAAGACGGGACTGACAGGCTGCACCTTCATGGCC 720
    |||
Db 768 GACTGGAAGAAGCGCTGTGCCATGCTGAAGACGGGACTGACAGGCTGCACCTTCATGGCC 827
    |||
QY 721 TCTCCGCACTGTGT 734
    |||
Db 828 TCTCCGCACTGTGT 841
    |||

RESULT 9
LOCUS ALI135434
DEFINITION DKFZp76200115_r1 762 (synonym: hmcl2) Homo sapiens cDNA clone
ACCESSION ALI135434
VERSION ALI135434.1 GI:6603621
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 719)
AUTHORS Ansorge, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE EST (Ansorge, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp76200115) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
    source
        1..719
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="DKFZp76200115"
            /tissue type="melanoma (Mewo cell line)"
            /dev stage="adult"
            /lab_host="DH10B"
            /clone_lib="762 (synonym: hmcl2)"
            /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN
Query Match 47.4%; Score 717.4; DB 1; Length 719;
Best Local Similarity 99.9%; Pred. No. 1.2e-187;
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 360 GGAAGGAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 419
    |||
Db 1 GGAAGGAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 60
    |||

QY 420 GCTTTTACAGGCATTTCTTCAGTGTCTTCTACAGAGAACATGGAGCCAGATTTTC 479
    |||
Db 61 GCTTTTACAGGCATTTCTTCAGTGTCTTCTACAGAGAACATGGAGCCAGATTTTC 120
    |||

QY 480 TCCAAAGAACATCGGTGCTTCCCTGCCAGGAGCCCGCCAGTTCCTCCCTCCTAGA 539
    |||
Db 121 TCCAAAGAACATCGGTGCTTCCCTGCCAGGAGCCCGCCAGTTCCTCCCTCCTAGA 180
    |||

QY 540 AGAACACAGGCCACTACAGAAAGTACATGGTATGCTCAGACCGAAATGGTGAAGACGGGAGA 599
    |||

```

```

Db 181 GGAACACAGGCCACTACAGAAATGATGGTATGCTCAGACGAAATGGTGAAGACGGGAGA 240
    |||
QY 600 GGGCCAGATTATGCCCACCTTTGTCCGGCCCTATGTGGCCATTCTCTGCGCATTTGGCTCT 659
    |||
Db 241 GGGCCAGATTATGCCCACCTTTGTCCGGCCCTATGTGGCCATTCTCTGCGCATTTGGCTCT 300
    |||
QY 660 TGACTGGAAGAAGCGCTGTGCCATGCTGAAGACGGGACTGACAGGCTGCACCTTCATGGC 719
    |||
Db 301 TGACTGGAAGAAGCGCTGTGCCATGCTGAAGACGGGACTGACAGGCTGCACCTTCATGGC 360
    |||
QY 720 CTCTCCGCACTGTGTGGCTTACAGCCGAGACAGCGGCCCTCCCTCAGTATGATGTG 779
    |||
Db 361 CTCTCCGCACTGTGTGGCTTACAGCCGAGACAGCGGCCCTCCCTCAGTATGATGTG 420
    |||
QY 780 CTTCCCTGACTGAAAGACAGATCCAGAGGCTGTCAGCTCTGGTGGTGGTGGTGGTGGTGG 839
    |||
Db 421 CTTCCCTGACTGAAAGACAGATCCAGAGGCTGTCAGCTCTGGTGGTGGTGGTGGTGGTGG 480
    |||
QY 840 CCAGTCGCTTACGTTGCTTACTGATTCGAGAGTTATGTGCTGAGCTTCCCAACAGCTCTT 899
    |||
Db 481 CCAGTCGCTTACGTTGCTTACTGATTCGAGAGTTATGTGCTGAGCTTCCCAACAGCTCTT 540
    |||
QY 900 CAAAGGAAGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 959
    |||
Db 541 CAAAGGAAGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 600
    |||
QY 960 CTTCCGCAAGCCGACCATTTATTTGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1019
    |||
Db 601 CTTCCGCAAGCCGACCATTTATTTGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
    |||

QY 1020 GCGGAGCGGACCTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1078
    |||
Db 661 GCGGAGCGGACCTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719
    |||

```

```

RESULT 10
LOCUS BI524031
DEFINITION 603052202P1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201616 5',
    mRNA sequence.
ACCESSION BI524031
VERSION BI524031.1 GI:15348823
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 889)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11505 row: d column: 01
High quality sequence stop: 880.
FEATURES
    source
        1..889
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5201616"
            /lab_host="DH10B"
            /clone_lib="NIH_MGC_122"
            /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                Site_1: NotI; Site_2: EcoRV (destroyed); RNA source"

```

anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 47.0%; Score 712.2; DB 4; Length 889;  
Best Local Similarity 99.1%; Pred. No. 3; 7e-106;  
Matches 758; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

Qy 1 ATGCCCGGGCTCTCGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCGC 60  
Db 87 ATGCCCGGGCTCTCGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCGC 146  
Qy 61 TTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Db 147 TTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 206  
Qy 121 ACCTTGGCTGTCTCTCTTGGATTGAGTACGAGCATCAAGGCTCTCTTTCACCAACCTC 180  
Db 207 ACCTTGGCTGTCTCTCTTGGATTGAGTACGAGCATCAAGGCTCTCTTTCACCAACCTC 266  
Qy 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATCGGTCATC 240  
Db 267 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATCGGTCATC 326  
Qy 241 AGCTTGGAGGATTCATGAGAGGCTGGCACCACCCACCTGCGCCCTGAGAGCGGGTG 300  
Db 327 AGCTTGGAGGATTCATGAGAGGCTGGCACCACCCACCTGCGCCCTGAGAGCGGGTG 386  
Qy 301 GCATCTCTTTGAGTGGCGCCGAGCCAGCCAGATGAAGACGTGCCCCATGAG 360  
Db 387 GCATCTCTTTGAGTGGCGCCGAGCCAGCCAGATGAAGACGTGCCCCATGAG 446  
Qy 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAGTGGAG 420  
Db 447 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAGTGGAG 506  
Qy 421 CTTTTCACGCAATTCCTTCACTGCTCTTACAGAGCAATGGAGCCAGAGATTTTCT 480  
Db 507 CTTTTCACGCAATTCCTTCACTGCTCTTACAGAGCAATGGAGCCAGAGATTTTCT 566  
Qy 481 CCAAGGAACATCCGGTGTCTTGGCCCTGCGAGGAGCCCGCCAGTTCGCCCGCTCTAGAA 540  
Db 567 CCAAGGAACATCCGGTGTCTTGGCCCTGCGAGGAGCCCGCCAGTTCGCCCGCTCTAGAG 626  
Qy 541 GAACACAGCCACTACAGAGTACATGTATGTGTCAGACGAATGTGAAGACGGAGAG 600  
Db 627 GAACACAGCCACTACAGAGTACATGTATGTGTCAGACGAATGTGAAGACGGAGAG 686  
Qy 601 GCCAGATTGATGCCACCTGTGCGGCC-CTATGTGGCATTCATCTGCGATTTGCTC 659  
Db 687 GCCAGATTGATGCCACCTGTGCGGCCCACTATGTGGCATTCATCTGCGATTTGCTC 746  
Qy 660 TGACT- GGAAGACGCTGTGCCATGCTGAGAGAGCGG- ACTGCGAGCTGCGACTTCATG 717  
Db 747 TGACTGGAAGACGCTGTGCCATGCTGAGAGAGCGGAACTGCGAGCTGCGACTTCATG 806  
Qy 718 GCCTCTCCGCACTGTGTTGGCTTACAGCCGACAGCAAGCGGCCCC 762  
Db 807 GCCTCTCCGCACTGTGTTGGCTTACAG-CCGACGACAGCGGCCCC 850

RESULT 11  
BX449172  
LOCUS  
DEFINITION BX449172 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
CS0DM014YK11 5-PRIME, mRNA sequence.  
ACCESSION BX449172

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BX449172.2 GI:47067286  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 22, 2003 this sequence version replaced gi:31030014.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
This sequence belongs to sequence cluster 9980.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0AM014AF06QPl&c=9980.f.  
Location/Qualifiers  
1..898  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM014YK11"  
/tissue\_type="FETAL LIVER"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL LIVER"  
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

FEATURES  
source

Query Match 46.7%; Score 706.6; DB 5; Length 898;  
Best Local Similarity 97.8%; Pred. No. 1.3e-184;  
Matches 744; Conservative 4; Mismatches 10; Indels 3; Gaps 3;

Qy 1 ATGCCCGGGCTCTCGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCGC 60  
Db 139 ATGCCCGGGCTCTCGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCGC 198  
Qy 61 TTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Db 199 TTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 258  
Qy 121 ACCTTGGCTGTCTCTCTTGGATTGAGTACGAGCATCAAGGCTCTCTTTCACCAACCTC 180  
Db 259 ACCTTGGCTGTCTCTCTTGGATTGAGTACGAGCATCAAGGCTCTCTTTCACCAACCTC 318  
Qy 181 CATGTGCTCTACGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
Db 319 CATGTGCTCTACGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 378  
Qy 241 ACCTTGGAGGATTTTCATGAGAGAGCTGGACCCACCCACTGCGCCCTGAGAGCGGGTG 300  
Db 379 AGCTTGGAGGATTTTCATGAGAGAGCTGGACCCACCCACTGCGCCCTGAGAGCGGGTG 438  
Qy 301 GCATCTCTTTGAGGTGGCAGCCCGAGAGCCAGATGAAGAGAGCTGCCCCCATGAGAG 360  
Db 439 GCATCTCTTTGAGGTGGCAGCCCGAGAG-CCAGATGAAGAGAGCTGCCCCCATGAGAG 497  
Qy 421 CTTTTCACGCAATTCCTTCACTGCTCTTACAGAGCAATGGAGCCAGAGATTTTCT 480

ORIGIN

```

Db      558 CTTTTCACAGGCAATTCCTTCAGTGTCTCGTACAGAGAA CAATGGAGCCAGAGATTTTCT 617
Qy      481 CCAAGGAAACATCCGGTGTCTGCCCTGCCAGGAGCCCCAGCCAGTTCGCCGTCTAGAA 540
Db      618 CCNAGGACATCCGGTGTCTGCCCTGCCAGGAGCCCCAGCCAGTTCGCCGTCTAGAG 677
Qy      541 GAACACAGGCCACTACAGAAAGTACATGGTATGGTTCAGAGC- AAATGGTGAAGACGGGAGA 599
Db      678 GAACACAGGCCACTACAGAAAGTACATGGTATGGTTCAGAGC- AAATGGTGAAGACGGGAGA 737
Qy      600 GGGCCAGATTCATGCCCACTGTTCGGGCCCTATGCGGCATTCATTCGGCATTTGGCTC 659
Db      738 CGCCAGATTCATGCCCACTGTTCGGGCCCTATGCGGCATTCATTCGGCATTTGGCTC 797
Qy      660 TGACTGGAAGAACCGCTGTGCGCATGTGAAGACGGGACTGCAGGCTCGCACTTCATGCG 719
Db      798 TGACTGGAAGAACCGCTGTGCGCATGTGAAGACGGGACTGCAGGCTCGCACTTCATGCG 857
Qy      720 CTCTCCCGAGTGTGGGCTTACAGCCGACGACACGCGCC 759
Db      858 CTCTCCCGAGTGTGGGCTTACAGCCGACGACACGCGCC 898

RESULT 12
BM475212
LOCUS      704 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT 6477457 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559347
5', mRNA sequence.
ACCESSION BM475212
VERSION    1
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 704)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12283 row: m column: 12
          High quality sequence stop: 685.
          Location/Qualifiers
            1..704
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:5559347"
              /tissue_type="duodenal adenocarcinoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_88"
              /note="Organ: small intestine; Vector: pCMV-SPORT6;
              Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
              oligo-dT primed. Average insert size 1.767 kb. Library
              enriched for full-length clones and constructed by Life
              Technologies. Note: this is a NIH_MGC Library."

FEATURES
             source
             1..704
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:5559347"
               /tissue_type="duodenal adenocarcinoma, cell line"
               /lab_host="DH10B (phage-resistant)"
               /clone_lib="NIH_MGC_88"
               /note="Organ: small intestine; Vector: pCMV-SPORT6;
               Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
               oligo-dT primed. Average insert size 1.767 kb. Library
               enriched for full-length clones and constructed by Life
               Technologies. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 45.2%; Score 683.8; DB 4; Length 704;
Best Local Similarity 99.7%; Pred. No. 2.7e-178;
Matches 685; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      797 AGATCAGAGGGCTGTGAAGCTCTCGGTGAGTTCGCTGAGTCCCGAGTCCCGTCTACGTTG 856
Db      3 AGATCAGAGGGCTGTGAAGCTCTCGGTGAGTTCGCTGAGTCCCGAGTCCCGTCTACGTTG 62

```

```

Qy      857 CTACTGATTTCCGAGAGTTATGTGCTGAGCTCAACAGAGCTCTTCAAAGGAAAGGTGAAG 916
Db      63 CTACTGATTTCCGAGAGTTATGTGCTGAGCTCAACAGAGCTCTTCAAAGGAAAGGTGAAG 122
Qy      917 TGGTGAAGCTGAAGCCTGAGGTGGCCAGGTGCAGCTGTACATCTCTCGGCCAAGCGGACC 976
Db      123 TGGTGAAGCCTGAAGCCTGAGGTGGCCAGGTGCAGCTGTACATCTCTCGGCCAAGCGGACC 182
Qy      977 ACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1036
Db      183 ACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 242
Qy      1037 AGGGAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1096
Db      243 AGGGAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 302
Qy      1097 GATTTCTGGCCGAGCACCAGACCCCTCTGATCTCTGGAGGACCCAGAGTCTGAGCTGTGCTCT 1156
Db      303 GATTTCTGGCCGAGCACCAGACCCCTCTGATCTCTGGAGGACCCAGAGTCTGAGCTGTGCTCT 362
Qy      1157 TCCAGCCAGGCTGCGCAGCAGAGTGTCTCGGGATTGCAAACTCTCTCTCTCTCTCTCTCTCTCT 1216
Db      363 TCCAGCCAGGCTGCGCAGCAGAGTGTCTCGGGATTGCAAACTCTCTCTCTCTCTCTCTCTCTCT 422
Qy      1217 AAAGATGAGAAAGAGTGTCCAGGAGCCCTCAAGGAGGAGAGAGCGCTCCATATCCAGGGCA 1276
Db      423 AAAGATGAGAAAGAGTGTCCAGGAGCCCTCAAGGAGGAGAGAGCGCTCCATATCCAGGGCA 482
Qy      1277 TAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTCTGTCTCTTC 1336
Db      483 TAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTCTGTCTCTTC 542
Qy      1337 TGGGAATTTCTCAGCTGCGCAAGCAGTCCAGCCTCGCTTCTCTGTCCACTCTGTCTGTG 1396
Db      543 TGGGAATTTCTCAGCTGCGCAAGCAGTCCAGCCTCGCTTCTCTGTCCACTCTGTCTGTG 602
Qy      1397 AGCAGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATTT 1456
Db      603 AGCAGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATTT 662
Qy      1457 TGATACAAAGGTGATGACTATCTAGAA 1483
Db      663 TGATACAAAGGTGATGACTATCTAGAA 689

RESULT 13
CO774833
LOCUS      942 bp mRNA linear EST 04-AUG-2004
DEFINITION ILLUMIGEN MCQ 51950 Katze MNLV Macaca nemestrina cDNA clone
          IBUM:29950 5', similar to Bases 7 to 852 highly similar to human
          POFU1 (Hs.178292), mRNA sequence.
ACCESSION CO774833
VERSION    1
KEYWORDS   EST.
SOURCE     Macaca nemestrina (pig-tailed macaque)
ORGANISM   Macaca nemestrina
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 942)
AUTHORS   Katze M.G., Thomas M., Korth M., Iadonato S.P. and Magness C.L.
TITLE     Large-scale Rhesus Macaque cDNA Sequencing
JOURNAL   Unpublished (2003)
COMMENT   Contact: C. Magness
          Illumigen Biosciences Inc.
          2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
          Tel: 2063780400
          Fax: 2063780408
          Email: cmagness@illumigen.com
          Sequenced on 2004.07.21. 611 Q20 bases. Library Preparation: Prof.
          Michael Katze Lab at University of Washington DNA Sequencing:
          Illumigen Biosciences Inc. For further information, see

```





Search completed: November 9, 2005, 10:38:05  
Job time : 5495.13 secs

---



Db 241 ACCTTGGAGGATTTTCATGGAGAACTGGGACCCACCACCTGGCCCCCTGAGAACGGGTG 300  
Qy\* 301 GCATACCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAGACGTGCCCCATGAAG 360  
Db 301 GCATACCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAGACGTGCCCCATGAAG 360  
Qy 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTCACAGCATTTCTCTTCAGTGTCTTCTACAGAGAAATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTTCACAGCATTTCTCTTCAGTGTCTTCTACAGAGAAATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAAACATCCGGTGTCTTGCCTGCGCAGAGCCCGCCAGTTCGCCCGCTAGAA 540  
Db 481 CCAAGGAAACATCCGGTGTCTTGCCTGCGCAGAGCCCGCCAGTTCGCCCGCTAGAA 540  
Qy 541 GAACACAGGCCACTACAGAGATACATGTATGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Db 541 GAACACAGGCCACTACAGAGATACATGTATGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Qy 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGATTCATCTGCGCATTTGGCTCT 660  
Db 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGATTCATCTGCGCATTTGGCTCT 660  
Qy 661 GACTGGAAGAACGCTGTGCGCATCTGTGAAGACGGGACTGCGGCTCGCACTTCATGSCC 720  
Db 661 GACTGGAAGAACGCTGTGCGCATCTGTGAAGACGGGACTGCGGCTCGCACTTCATGSCC 720  
Qy 721 TCTCCGCAAGTGTGGGCTACAGCCGACACAGCGCCCGCCCTCACGATGACTATGTC 780  
Db 721 TCTCCGCAAGTGTGGGCTACAGCCGACACAGCGCCCGCCCTCACGATGACTATGTC 780  
Qy 781 CTGCTGACCTGAAGGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
Db 781 CTGCTGACCTGAAGGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
Qy 841 CAGTCCGTCTACGTGTCTACTGATTCGAGAGTTATGTGCTCAGCTCCCAACAGCTCTTC 900  
Db 841 CAGTCCGTCTACGTGTCTACTGATTCGAGAGTTATGTGCTCAGCTCCCAACAGCTCTTC 900  
Qy 901 AAAGGGAAGGTGAAGGTGGTGAAGCTGAGGCTGAGGTGGCCAGCTCCACCTGTACATC 960  
Db 901 AAAGGGAAGGTGAAGGTGGTGAAGCTGAGGCTGAGGTGGCCAGCTCCACCTGTACATC 960  
Qy 961 CTGCGCAAGCCGACCACTTTATGGCAACTGTGTCTCTCTCACTGCTTGTGGAAG 1020  
Db 961 CTGCGCAAGCCGACCACTTTATGGCAACTGTGTCTCTCTCACTGCTTGTGGAAG 1020  
Qy 1021 CGGAGCGGAGCTCTCAGGGAGGCGCTCTCTTTCTTGGGATGAGACAGGCCCGCTTAAG 1080  
Db 1021 CGGAGCGGAGCTCTCAGGGAGGCGCTCTCTTTCTTGGGATGAGACAGGCCCGCTTAAG 1080  
Qy 1081 CTGCGGAGCAGATTTCTGATT 1100  
Db 1081 CTGCGGAGCAGATTTCTGATT 1100

RESULT 2  
BD103196  
LOCUS BD103196 1100 bp DNA linear PAT 27-AUG-2002  
DEFINITION O-fucosyltransferase.  
ACCESSION BD103196  
VERSION BD103196.1 GI:22648770  
KEYWORDS JP 2001527389-A/9.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 1100)  
AUTHORS Wang, Y. and Spellman, M. W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: JP 2001527389-A 9 25-DEC-2001;

GENENTECH INC  
OS Unidentified  
PN JP 2001527389-A/9  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877  
PR WANG, MICHAEL W SPELLMAN 08/752498, 26-NOV-1997 US 08/978741 PI  
PC C12N15/54, C12N9/10, C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
CC Key Location/Qualifiers  
FT source 1..1100  
FT /organism='Unidentified'.  
FEATURES  
source  
1..1100  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
ORIGIN  
Query Match 100.0%; Score 1100; DB 6; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 1e-253;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGCCCGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCTGCGATGGGGCG 60  
Db 1 ATGCCCGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCTGCGATGGGGCG 60  
Qy 61 TTGGGAACACGAGCCGATCACTTTCTTGGGCTCTCTGGGATTTGCAAGCTGTAACCGT 120  
Db 61 TTGGGAACACGAGCCGATCACTTTCTTGGGCTCTCTGGGATTTGCAAGCTGTAACCGT 120  
Qy 121 ACCTTGGGTGTCCTCTCTGGATTTGAGTACAGCATCAAGCTCTCTTTCACCAACCTC 180  
Db 121 ACCTTGGGTGTCCTCTCTGGATTTGAGTACAGCATCAAGCTCTCTTTCACCAACCTC 180  
Qy 181 CATGTGCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
Db 181 CATGTGCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
Qy 241 ACCTTGGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGCCCTCGAAGAGCGGGTG 300  
Db 241 ACCTTGGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGCCCTCGAAGAGCGGGTG 300  
Qy 301 GCATACCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAGACGTGCCCCATGAAG 360  
Db 301 GCATACCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAGACGTGCCCCATGAAG 360  
Qy 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTCACAGCATTTCTCTTCAGTGTCTTCTACAGAGAAATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTTCACAGCATTTCTCTTCAGTGTCTTCTACAGAGAAATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAAACATCCGGTGTCTTGCCTGCGCAGAGCCCGCCAGTTCGCCCGCTAGAA 540  
Db 481 CCAAGGAAACATCCGGTGTCTTGCCTGCGCAGAGCCCGCCAGTTCGCCCGCTAGAA 540  
Qy 541 GAACACAGGCCACTACAGAGATACATGTATGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Db 541 GAACACAGGCCACTACAGAGATACATGTATGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Qy 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGATTCATCTGCGCATTTGGCTCT 660  
Db 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGATTCATCTGCGCATTTGGCTCT 660  
Qy 661 GACTGGAAGAACGCTGTGCGCATCTGTGAAGACGGGACTGCGGCTCGCACTTCATGSCC 720  
Db 661 GACTGGAAGAACGCTGTGCGCATCTGTGAAGACGGGACTGCGGCTCGCACTTCATGSCC 720



```
QY 721 TCTCCGACGTGTGTGGCTTACAGCCGACAGCAGCGGCCCCCTCACAGATGACTATGTGC 780
Db 721 TCTCCGACGTGTGTGGCTTACAGCCGACAGCAGCGGCCCCCTCACAGATGACTATGTGC 780
QY 781 CTGCGCTGACCTGAAGGAGATCAGAGGGCTGTGAAGCTCTGGTGAGGTTCGTGATGCC 840
Db 781 CTGCGCTGACCTGAAGGAGATCAGAGGGCTGTGAAGCTCTGGTGAGGTTCGTGATGCC 840
QY 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTTCCAGAGCTTCCAAAGCTCTTC 900
Db 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTTCCAGAGCTTCCAAAGCTCTTC 900
QY 901 AAAGGAAAGGTGAAGTGTGAGCCTGAAGCCTGAGGTGGCCAGCAGTTCGACCTGTACATC 960
Db 901 AAAGGAAAGGTGAAGTGTGAGCCTGAAGCCTGAGGTGGCCAGCAGTTCGACCTGTACATC 960
QY 961 CTGCGCCAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTGTGAAG 1020
Db 961 CTGCGCCAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTGTGAAG 1020
QY 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTCTTCCGATGGAAGGCGCCCTAAG 1080
Db 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTCTTCCGATGGAAGGCGCCCTAAG 1080
QY 1081 CTGCGGACGAGTTCGTATT 1100
Db 1081 CTGCGGACGAGTTCGTATT 1100

RESULT 3
LOCUS AR163451 1514 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 2 from patent US 6270987.
ACCESSION AR163451
VERSION AR163451.1 GI:16234056
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 2 07-AUG-2001;
FEATURES
source
1..1514
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 1100; DB 6; Length 1514;
Best Local Similarity 100.0%; Pred. No. 1e-253;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCGGGGTCTCTGGAGCCCGCGGGTTACCTGCTCTACTGCCCTGATGGGGCGC 60
Db 1 ATGCGCGGGGTCTCTGGAGCCCGCGGGTTACCTGCTCTACTGCCCTGATGGGGCGC 60
QY 61 TTTGGGAACCGGCGGATCTTCTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 120
Db 61 TTTGGGAACCGGCGGATCTTCTGGGCTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 120
QY 121 ACCTTGGGTGTCCTCTTGGATTGAGTACAGCATCAAGCCTCTCTTTCACCAACCTC 180
Db 121 ACCTTGGGTGTCCTCTTGGATTGAGTACAGCATCAAGCCTCTCTTTCACCAACCTC 180
QY 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCAATC 240
Db 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCAATC 240
QY 241 AGCTTGGAGGATTTTCATGGAGAGCTGGACCCACCACTGGCCCTGAGAGCGGTG 300
Db 241 AGCTTGGAGGATTTTCATGGAGAGCTGGACCCACCACTGGCCCTGAGAGCGGTG 300
```

```
QY 301 GCATATCTGCTTTAGGTGGCCAGCCAGCGAAGCCAGATAAGAACGTCGCCCAATGAAG 360
Db 301 GCATATCTGCTTTAGGTGGCCAGCCAGCGAAGCCAGATAAGAACGTCGCCCAATGAAG 360
QY 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGATTTCAACAGTCGAG 420
Db 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGATTTCAACAGTCGAG 420
QY 421 CTTTTCACAGGATTTCTTTCAGTCTTCTTACAGAACATGTCGACCAATGAGCCAGAGATTTCT 480
Db 421 CTTTTCACAGGATTTCTTTCAGTCTTCTTACAGAACATGTCGACCAATGAGCCAGAGATTTCT 480
QY 481 CCAAGGAAACATCCGGTGTCTTCCCTGCGAGGAGCCAGCCAGTTCGCCCTCTCTAGAA 540
Db 481 CCAAGGAAACATCCGGTGTCTTCCCTGCGAGGAGCCAGCCAGTTCGCCCTCTCTAGAA 540
QY 541 GAAACACAGGCCCATACAGAAATGATGTATGTCAGACGAAATGTCGAGGAGCGGAGAG 600
Db 541 GAAACACAGGCCCATACAGAAATGATGTATGTCAGACGAAATGTCGAGGAGCGGAGAG 600
QY 601 GCCCAGATTTTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 660
Db 601 GCCCAGATTTTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 660
QY 661 GACTGGAAGAACCGCTGTGCCATGCTGAAGGACGGGACTGCAAGGCTCGCACTTCATGGCC 720
Db 661 GACTGGAAGAACCGCTGTGCCATGCTGAAGGACGGGACTGCAAGGCTCGCACTTCATGGCC 720
QY 721 TCTCCGAGTGTGTGGGCTACAGCGCAGCAGCGGCCCTCACGATGACTATGTGC 780
Db 721 TCTCCGAGTGTGTGGGCTACAGCGCAGCAGCGGCCCTCACGATGACTATGTGC 780
QY 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTTCGCTGATGCC 840
Db 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTTCGCTGATGCC 840
QY 841 CAGTCGGTCTACGTTGCTACTGATTTCCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900
Db 841 CAGTCGGTCTACGTTGCTACTGATTTCCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900
QY 901 AAAGGAAAGGTGAAGTGTGAGCCTGAAGCCTGAGGTGGCCAGCAGTTCGATACATC 960
Db 901 AAAGGAAAGGTGAAGTGTGAGCCTGAAGCCTGAGGTGGCCAGCAGTTCGATACATC 960
QY 961 CTGCGCCAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTGTGAAG 1020
Db 961 CTGCGCCAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTGTGAAG 1020
QY 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTCTTCCGATGGAAGGCGCCCTAAG 1080
Db 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTCTTCCGATGGAAGGCGCCCTAAG 1080
QY 1081 CTGCGGACGAGTTCGTATT 1100
Db 1081 CTGCGGACGAGTTCGTATT 1100

RESULT 4
LOCUS BD103188 1514 bp DNA linear PAT 27-AUG-2002
DEFINITION O-fucosyltransferase.
ACCESSION BD103188
VERSION BD103188.1 GI:22648762
KEYWORDS JP 2001527389-A/1.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1514)
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 1 25-DEC-2001;
COMMENT GENENTECH INC
OS Unidentified
```

|   |   |     |   |  |      |
|---|---|-----|---|--|------|
| D   | b |     | 721   | TCTCCGCACTGTGTGGGCTACAGCCGACACAGGGGCCCCCCTCACGATGACTATGTGC       | 780  |
| Q   | y |     | 781   | CTGCCTGAAGCTCAGAGAGATCAGAGGGCTGTGAAGCTCTTGGTGAGGTGCGTGGATGCC     | 840  |
| D   | b |     | 781   | CTGCCTGAAGCTCAGAGAGATCAGAGGGCTGTGAAGCTCTTGGGTGAGGTGCGTGGATGCC    | 840  |
| Q   | y |     | 841   | CAGTCGGTCTACGTTGCTACTGATTCCGAGAGATTATGTGCTCAGCTCCAAGCTCTTC       | 900  |
| D   | b |     | 841   | CAGTCGGTCTACGTTGCTACTGATTCCGAGAGATTATGTGCTCAGCTCCAAGCTCTTC       | 900  |
| Q   | y |     | 901   | AAAGGAAAGGTGAAGGTGGTGAGCCCTGAAGCCCTGAGGTGGCCCCAGGCTGCACCTGTACATC | 960  |
| D   | b |     | 901   | AAAGGAAAGGTGAAGGTGGTGAGCCCTGAAGCCCTGAGGTGGCCCCAGGCTGCACCTGTACATC | 960  |
| Q   | y |     | 961   | CTCGGCCAAGCCGACCACTTATTGTGCAACTGTGTCTCTTCACTGCCCTTTGTGAAG        | 1020 |
| D   | b |     | 961   | CTCGGCCAAGCCGACCACTTATTGTGCAACTGTGTCTCTTCACTGCCCTTTGTGAAG        | 1020 |
| Q   | y |     | 1021  | CGGAGCGGACCTCCAGGGGAGGCGCTCTCTTCTTTCGGCATGACAAGGCCCCCTTAAG       | 1080 |
| D   | b |     | 1021  | CGGAGCGGACCTCCAGGGGAGGCGCTCTCTTCTTTCGGCATGACAAGGCCCCCTTAAG       | 1080 |
| Q   | y |     | 1081  | CTCGGGACGAGTCTTGATT  | 1100 |
| D   | b |     | 1081  | CTCGGGACGAGTCTTGATT  | 1100 |
| RESULT 5  |   |     |   |  |      |
| AR163452  |   |     |   |  |      |
| LOCUS AR163452 1300 bp DNA linear PAT 17-OCT-2001             |   |     |   |  |      |
| DEFINITION Sequence 6 from patent US 6270987.                 |   |     |   |  |      |
| ACCESSION AR163452  |   |     |   |  |      |
| VERSION AR163452.1 GI:16234057                                |   |     |   |  |      |
| KEYWORDS Unknown.   |   |     |   |  |      |
| SOURCE Unclonified.   |   |     |   |  |      |
| ORGANISM O-fucosyltransferase                                 |   |     |   |  |      |
| REFERENCE 1 (bases 1 to 1300)                                 |   |     |   |  |      |
| AUTHORS wang,Y. and Spellman,M.W.                             |   |     |   |  |      |
| TITLE O-fucosyltransferase                                    |   |     |   |  |      |
| JOURNAL Patent: US 6270987-A 6 07-AUG-2001;                   |   |     |   |  |      |
| FEATURES Location/Qualifiers                                  |   |     |   |  |      |
| source 1..1300  |   |     |   |  |      |
| /organism="unknown"   |   |     |   |  |      |
| /mol_type="unassigned DNA"                                    |   |     |   |  |      |
| ORIGIN  |   |     |   |  |      |
| Query Match 99.9%; Score 1098.4; DB 6; Length 1300;           |   |     |   |  |      |
| Best Local Similarity 99.9%; Pred. No. 2.4e-253;              |   |     |   |  |      |
| Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0; |   |     |   |  |      |
| Q   | y | 1   | ATGCCCGCGGGCTCTGTGGACCCGCGGGGTACTGCTACTCTACTGCCCTGCATGGGGCGC    | 60   |      |
| D   | b | 136 | ATGCCCGCGGGCTCTGTGGACCCGCGGGGTACTGCTACTCTACTGCCCTGCATGGGGCGC    | 195  |      |
| Q   | y | 61  | TTTGGGNACAGGCCGATCACTTTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT        | 120  |      |
| D   | b | 196 | TTTGGGNACAGGCCGATCACTTTTGGGCTCTCTGGGCTTTTGCNAAGCTGTAAACCGT      | 255  |      |
| Q   | y | 121 | ACCTTGGCTGTCCCTCTTGTGATGTAGTACCAGATCAAGACCTCTTTACGACCTTC        | 180  |      |
| D   | b | 256 | ACCTTGGCTGTCCCTCTTGTGATGTAGTACCAGATCAAGACCTCTTTACGACCTTC        | 315  |      |
| Q   | y | 181 | CATGTGTCCTACAGAAGTACTTCAAGCTGTGAGGCCCTCTCAGGCTTACCATCGGGTCATC   | 240  |      |
| D   | b | 316 | CATGTGTCCTACAGAAGTACTTCAAGCTGTGAGGCCCTCTCAGGCTTACCATCGGGTCATC   | 375  |      |
| Q   | y | 241 | AGCTTGGAGGATTTTCAATGGAGAGCTGTGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG  | 300  |      |
| D   | b | 376 | AGCTTGGAGGATTTTCAATGGAGAGCTGTGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG  | 435  |      |
| Q   | y | 301 | GCATCTAGCTTTTGTGGGTGGCACGCCAGCGAAGCCACAGATAGAGAAGCTGCCCCCATGAAG | 360  |      |







The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1

JOURNAL  
MEDLINE  
PUBMED  
PUBMED  
REFERENCE  
AUTHORS

Chiang, P.W., Wang, S., Smith, P., Song, W.J., Ramamoorthy, S., Hillman, J., Puett, S., Van Keuren, M.L., Crombez, E., Kumar, A., Glover, T.W., Miller, D.E., Tsai, C.H., Blackburn, C.C., Chen, X.N., Sun, Z., Cheng, J.F., Kornberg, J.R., and Kurnit, D.M.  
Identification and analysis of the human and murine putative chromatin structure regulator SUP6H and Supt6h  
Genomics 34 (3), 328-333 (1996)

TITLE

JOURNAL  
MEDLINE  
PUBMED  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.  
Direct Submission  
Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)  
On May 9, 2002 this sequence version replaced gi:1136419.

COMMENT

FEATURES  
source

1. 5189  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="ha02567s1"  
/sex="male"  
/cell\_line="KG-1"  
/cell\_type="myeloblast"  
/tissue\_type="brain"  
/clone\_lib="pBluescriptII SK plus"  
/note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1. 1066 was derived from p100933 and 1067 - 5189 was derived from ha02567)."  
1. 5189  
/gene="KIAA0180"  
1. 1213  
/gene="KIAA0180"  
/citation=[2]  
/codon\_start=2  
/protein\_id="BAA11497.2"  
/db\_xref="GI:20521838"  
/translation="PPRLCAAGSGRADMGAAWAPLVSFLLLPLFGMPAGSW DPAGLLYCPGMRGNQADHFLGSLAPAKLLNRTLAVPPWIEYQHHKPPPTNLHVSY QKYPLEPLQAVHRVTSLEDFMEKLPATHWPPEKRVAYCFEVAQRSPDKKTCPMKEG NPFGPFDQPHYSFNKSELTGIGSFSASYREWSQSFSPKEHPVLALPGAPQFPVLE BHRPLOKYMWSDEMVKTEAQIHAHLVVRPVVGIHLRIGSDWKNACMLKDPGTAGSHF MASPCQGYGSRSTAPLMTMCLPDKEIORAVKLWVRSIDQSYVYVATDSYVPEL QQLPKGVKVSILKPEVAQVDLYILGQADHFIGNCVSFTAFVKRERDLQGRPSFPG MDRPPLKLDEF"

gene

CDS

ORIGIN

Query Match 99.78; Score 1096.8; DB 9; Length 5189;  
Best Local Similarity 99.8%; Pred. No. 5.9e-253; Indels 0; Gaps 0;  
Matches 1098; Conservative 0; Mismatches 2;

Qy 1 ATGCCCGGGCTCTGGGACCCGGCGGTACTGCTCTACTGCCCCCTGCGATGGGGCGC 60  
Db 116 ATGCCCTGGGGCTCTGGGACCCGGCGGTACTGCTCTACTGCCCCCTGCGATGGGGCGC 175  
Qy 61 TTGGGAACACGAGCGCATCTTTGGGCTCTCTGGCATTTGCAAGCTGTCAACCGT 120  
Db 176 TTGGGAACACGAGCGCATCTTTGGGCTCTCTGGCATTTGCAAGCTGTCAACCGT 235  
Qy 121 ACCTTGGGTGTCCTCTGGATTTAGTACACGATCACAAGCTCTCTTCCACCAACCTC 180  
Db 236 ACCTTGGGTGTCCTCTGGATTTAGTACACGATCACAAGCTCTCTTCCACCAACCTC 295  
Qy 181 CATGTGTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTGTCATC 240

Db 296 CATGTGTCTACCAAGAGTACTTCAAGCTGGAGCCCTCAGGCTTACCATCGGTGTCATC 355  
Qy 241 AGCTTGGAGGATTTTCATGGAGAAGCTGGCAACCCACCTGGCCCCCTCGAGAAGCGGTG 300  
Db 356 AGCTTGGAGGATTTTCATGGAGAAGCTGGCAACCCACCTGGCCCCCTCGAGAAGCGGTG 415  
Qy 301 GCATATGCTTTGAGGTGGCGCCAGCGAAGCCAGATAGAAAGACGTGCCCCCATGAAG 360  
Db 416 GCATATGCTTTGAGGTGGCGCCAGCGAAGCCAGATAGAAAGACGTGCCCCCATGAAG 475  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTAGTTTCAACAAGTCGAG 420  
Db 476 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTAGTTTCAACAAGTCGAG 535  
Qy 421 CTTTTTACAGGCATTTCTTCAGTCTCTCTACAGAAACAATGGAGCCAGAGATTTTCT 480  
Db 536 CTTTTTACAGGCATTTCTTCAGTCTCTCTACAGAAACAATGGAGCCAGAGATTTTCT 595  
Qy 481 CCAAGGAACAATCGGTGCTTGGCTTCCAGGAGCCCGCCAGCCAGCTTCCCTCTAGAA 540  
Db 596 CCAAGGAACAATCGGTGCTTGGCTTCCAGGAGCCCGCCAGCTTCCCTCTAGAG 655  
Qy 541 GAACACAGGCCCTTACAGAAAGTACATGTTGTTGTCAGACGAAATGGTGAAGACGGAG 600  
Db 656 GAACACAGGCCCTTACAGAAAGTACATGTTGTTGTCAGACGAAATGGTGAAGACGGAG 715  
Qy 601 GCCCAGATTCATGCCCCACCTTTGTCGGGCCCTATGTGGGCATTCATCTGGGCATTTGGCTCT 660  
Db 716 GCCCAGATTCATGCCCCACCTTTGTCGGGCCCTATGTGGGCATTCATCTGGGCATTTGGCTCT 775  
Qy 661 GACTGGAAGAACGCTTGTGCTCATGTCTGAAGGACGGGACTCGAGGCTCGCACTTTCATGGCC 720  
Db 776 GACTGGAAGAACGCTTGTGCTCATGTCTGAAGGACGGGACTCGAGGCTCGCACTTTCATGGCC 835  
Qy 721 TCTCCGCGAGTGTGGGCTACAGCGCGAGCACAGCGGCCCTCCTCAGATGACTATGTC 780  
Db 836 TCTCCGCGAGTGTGGGCTACAGCGCGAGCACAGCGGCCCTCCTCAGATGACTATGTC 895  
Qy 781 CTGCTGACCTGAAGAGGATCAGAGGCTGTGAAGCTCTGGGTGAGGTGGGTGATGCC 840  
Db 896 CTGCTGACCTGAAGAGGATCAGAGGCTGTGAAGCTCTGGGTGAGGTGGGTGATGCC 955  
Qy 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900  
Db 956 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 1015  
Qy 901 AAAGGAAGGTGAAGGTGGTGGAGCCCTGAAGCCTGAGGTGGCCCGACCTGTACATC 960  
Db 1016 AAAGGAAGGTGAAGGTGGTGGAGCCCTGAAGCCTGAGGTGGCCCGACCTGTACATC 1075  
Qy 961 CTCGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 1076 CTCGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1135  
Qy 1021 CGGGAGCGGAGACTTCCAGGGGAGCGCTTCTTTCTTTCGGCATGGACAGGCCCTTAAG 1080  
Db 1136 CGGGAGCGGAGACTTCCAGGGGAGCGCTTCTTTCTTTCGGCATGGACAGGCCCTTAAG 1195  
Qy 1081 CTGGCGGACGAGTTCTGATT 1100  
Db 1196 CTGGCGGACGAGTTCTGATT 1215

RESULT 10

Q0727777  
LOCUS  
DEFINITION  
ACCESSION  
Q0727777.1  
VERSION  
Q0727777.1  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Kits, such as nucleic acid arrays, comprising a majority of human exons or transcripts, for detecting expression and other uses thereof  
JOURNAL Patent: WO 02068579-A 13711 06-SEP-2002;  
PE Corporation (NY) (US)  
FEATURES Location/Qualifiers  
source 1..1167  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 99.7%; Score 1096.4; DB 6; Length 1167;  
Best Local Similarity 99.9%; Pred. No. 7.3e-253;  
Matches 1097; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGCCCGGGCTCTCTGGACCGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGGC 60  
DB 70 ATGCCCGGGCTCTCTGGACCGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGGC 129  
QY 61 TTGGGAACAGCGCGATCACTTTCTGGGCTCTCTGGGCAATTTGCAAGCTGCTAAACCGT 120  
DB 130 TTGGGAACAGCGCGATCACTTTCTGGGCTCTCTGGGCAATTTGCAAGCTGCTAAACCGT 189  
QY 121 ACCTTGGCTGCTCCCTCTCTGGATTGAGTACCAAGCATCAAGCCCTCTTACCAACCTC 180  
DB 190 ACCTTGGCTGCTCCCTCTCTGGATTGAGTACCAAGCATCAAGCCCTCTTACCAACCTC 249  
QY 181 CATGTGCTCTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
DB 250 CATGTGCTCTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 309  
QY 241 ACCTTGGAGGATTTATGAGAACTGCGACCCACCCACTGCGCCCTCTGAGAAGCGGTG 300  
DB 310 ACCTTGGAGGATTTATGAGAACTGCGACCCACCCACTGCGCCCTCTGAGAAGCGGTG 369  
QY 301 GCATCTGCTTTGAGTGGAGCGAGCGAGCCAGATGAGAGCGTCCCAATGAAG 360  
DB 370 GCATCTGCTTTGAGTGGAGCGAGCGAGCCAGATGAGAGCGTCCCAATGAAG 429  
QY 361 GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTCTGAGTTTCAAGTCGGAG 420  
DB 430 GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTCTGAGTTTCAAGTCGGAG 489  
QY 421 CTTTATCAGGCAATTTCTTCAAGTCTTCTACAGAGAACTATGAGCGCAGAGATTTCT 480  
DB 490 CTTTATCAGGCAATTTCTTCAAGTCTTCTACAGAGAACTATGAGCGCAGAGATTTCT 549  
QY 481 CCAAGGAACATCCGCTGCTGCTCCCTGCGAGAGCCCGAGCCAGTTCCTCGCTCTAGAA 540  
DB 550 CCAAGGAACATCCGCTGCTGCTCCCTGCGAGAGCCCGAGCCAGTTCCTCGCTCTAGAG 609  
QY 541 GAAACAGGCCACTACAGAGTACATGTTGCTGAGAGAACTGCTGAGAGCGGAGAG 600  
DB 610 GAAACAGGCCACTACAGAGTACATGTTGCTGAGAGAACTGCTGAGAGCGGAGAG 669  
QY 601 GCCAGATTCATGCCCACTTGTCCGGCCCTATGTTGGGCAATTCATCTCGCATTTGGCTCT 660  
DB 670 GCCAGATTCATGCCCACTTGTCCGGCCCTATGTTGGGCAATTCATCTCGCATTTGGCTCT 729  
QY 661 GACTGGAGAGCCCTGTGCTGCTGCTGAGAGCGGAGCTGAGGCTCCGCTCTCATGCCC 720  
DB 730 GACTGGAGAGCCCTGTGCTGCTGCTGAGAGCGGAGCTGAGGCTCCGCTCTCATGCCC 789  
QY 721 TCTCCGAGTGTGGGCTACAGCGCGAGCAGCGGCGCCCTCAGCATGACTATGTC 780  
DB 790 TCTCCGAGTGTGGGCTACAGCGCGAGCAGCGGCGCCCTCAGCATGACTATGTC 849  
QY 781 CTGCTGACCTGAAGAGGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGATGCC 840

DB 850 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGCTGATGCC 909  
QY 841 CAGTCGGTCTAGCTTGTCTACTGATTCGAGAGTATGTGCTGAGCTCCAAAGCTCTTC 900  
DB 910 CAGTCGGTCTAGCTTGTCTACTGATTCGAGAGTATGTGCTGAGCTCCAAAGCTCTTC 969  
QY 901 AAAGGAAAGGTGAAGGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGAGCTGTATCATC 960  
DB 970 AAAGGAAAGGTGAAGGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGAGCTGTATCATC 1029  
QY 961 CTGCGCCAAAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGGAAG 1020  
DB 1030 CTGCGCCAAAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGGAAG 1089  
QY 1021 CGGAGCGGAGCTCCAGGGAGGCGCTTCTTCTTCTGCGCATGAGAGCCCTTAAG 1080  
DB 1090 CGGAGCGGAGCTCCAGGGAGGCGCTTCTTCTTCTGCGCATGAGAGCCCTTAAG 1149  
QY 1081 CTGCGGAGCGAGTCTCTGA 1098  
DB 1150 CTGCGGAGCGAGTCTCTGA 1167  
RESULT 11  
AX780153  
LOCUS AX780153 5266 bp DNA linear PAT 14-JUL-2003  
DEFINITION Sequence 2310 from Patent WO03039443.  
ACCESSION AX780153  
VERSION AX780153.1 GI:32697147  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Haerlach, T., Schoch, C., Kern, W., Kohlmann, A., Schnittger, S.,  
Dugas, M., Ellis, R., Brors, B. and Mergenthaler, S.  
TITLE Novel genetic markers for leukemias  
JOURNAL Patent: WO 03039443-A 2310 15-MAY-2003;  
Deutsches Krebsforschungszentrum (DE);  
Ludwig-Maximilians-Universitaet Muenchen (DE);  
PD Dr. (DB); Schoch, Claudia (DB); Kern, Wolfgang (DE)  
FEATURES  
source 1..5266  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 99.6%; Score 1095.8; DB 6; Length 5266;  
Best Local Similarity 99.7%; Pred. No. 1e-252;  
Matches 1097; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGCCCGGGGCTCTCTGGACCCCGGCTTACCTGCTCTACTGCCCCCTGCATGGGGGC 60  
DB 132 ATGCTGCGGGCTCTCTGGACCCCGGCTTACCTGCTCTACTGCCCCCTGCATGGGGGC 191  
QY 61 TTGGGAACAGCGCGATCACTTTCTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 120  
DB 192 TTGGGAACAGCGCGATCACTTTCTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 251  
QY 121 ACCTTGGCTGCTCCCTCTCTGGATTTGAGTACAGATCAAGAGCTCTTTCACCAACCTC 180  
DB 252 ACCTTGGCTGCTCCCTCTCTGGATTTGAGTACAGATCAAGAGCTCTTTCACCAACCTC 311  
QY 181 CATGTGCTCTACAGAACTTCTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
DB 312 CATGTGCTCTACAGAACTTCTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 371  
QY 241 AGCTTGGAGGATTTATGAGAGAGCTGGCAACCCACTGCGCCCTCTGAGAAGCGGTG 300  
DB 372 AGCTTGGAGGATTTATGAGAGAGCTGGCAACCCACTGCGCCCTCTGAGAAGCGGTG 431





|||||  
711 GACTGGAAGAACGCTGTGCCATGCTGAAGGACGGGACTGCGAGGCTCGCACTTCATGGCC 770  
|||||  
721 TCTCCGCGAGTGTGGCTACAGCCGACAGAGCGGCCCCCTCACGATGATATGTC 780  
|||||  
771 TCTCCGCGAGTGTGGCTACAGCCGACAGAGCGGCCCCCTCACGATGATATGTC 830  
|||||  
781 CTGCTGACCTGAAGAGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGC 840  
|||||  
831 CTGCTGACCTGAAGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGC 890  
|||||  
841 CAGTCGGTCTAGCTGCTACTGATTCGAGAGTATGTCCTGAGCTCCAAAGCTCTTC 900  
|||||  
891 CAGTCGGTCTAGCTGCTACTGATTCGAGAGTATGTCCTGAGCTCCAAAGCTCTTC 950  
|||||  
901 AAAGGGAAGTGAAGGTGTGAGCTGAAGCTGAGGTGGCCGACGCTGACCTGTACATC 960  
|||||  
951 AAAGGGAAGTGAAGGTGTGAGCTGAAGCTGAGGTGGCCGACGCTGACCTGTACATC 1010  
|||||  
961 CTGCGCCAAAGCCGACACTTTATTTGGCACTGTGTCTCTCTCACTGCCCTTTGTGAAG 1020  
|||||  
1011 CTGCGCCAAAGCCGACACTTTATTTGGCACTGTGTCTCTCTCACTGCCCTTTGTGAAG 1070  
|||||  
1021 CCGGAGCGGAGCTCAGGGGAGGCGTCTTTCTTCGCGATGAGAGGCCCTTAAG 1080  
|||||  
1071 CCGGAGCGGAGCTCAGGGGAGGCGTCTTTCTTCGCGATGAGAGGCCCTTAAG 1130  
|||||  
1081 CTGCGGAGGAGTCTTGATT 1100  
|||||  
1131 CTGCGGAGGAGTCTTGATT 1150

RESULT 13  
AJ781500  
LOCUS  
DEFINITION  
AJ781500  
VERSION  
AJ781500.1 GI:50057069  
KEYWORDS  
fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.  
SOURCE  
Pan troglodytes (chimpanzee)  
ORGANISM  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
REFERENCE  
1 Martinez-Duncker, I., Mollicone, R., Candellier, J.J., Breton, C. and Oriol, R.  
TITLE  
A new superfamily of protein-O-fucosyltransferases, alpha2-fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs  
JOURNAL  
Glycobiology 13 (12), 1C-5C (2003)  
PUBMED  
12966037  
REFERENCE  
2 Martinez-Duncker, I., Oriol, R. and Mollicone, R.  
AUTHORS  
Phylogeny of fucosyltransferases  
TITLE  
Unpublished  
JOURNAL  
Unpublished  
REFERENCE  
3 (bases 1 to 1167)  
AUTHORS  
Oriol, R.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul  
Vaillant-Couturier, 94807, FRANCE  
FEATURES  
Location/Qualifiers  
1..1167  
/organism="Pan troglodytes"  
/mol\_type="mRNA"  
/db\_xref="taxon:9598"  
1..1167  
/gene="fut12"  
1..1167  
/gene="fut12"  
/function="adds fucose to Ser or Thr"  
/note="pofut1"  
/codon\_start=1

/product="protein-O-fucosyltransferase 1"  
/protein\_id="CAH03712.1"  
/db\_xref="GI:50057070"  
/translation="MGAAAWARPLSVSFLLLLPLPMPAGSDWDPAGVLLYLPCPMGRF  
GNQADHFLGSLAFKLRLNLTAVPEWIOHKKPFTNLHVSQYFKLEPQAHYRV  
ISLEDFMEKLAFTHPPEKRVAYCFEVAQORSPDKTCPMKEGNFPFWDQFHSFN  
KSELTGTSFSAVREOMSORFSPEKHPVLALPGAPQFPVLEHRPLQKYMVMWDEA  
VKTGBAQIHAHLVRPYGVGHLRIGSDKNACAMLDKGTAGSHFMAPOCVGVSRSTAA  
PLTWTMCLPDLKEIOLRAVKLWVRSIDAQSVVATDSYVPELOQLFKGKVKVSLKP  
EVAQVDLILQADHFIQNCVSSFTAFVRRERDLQGRSSFFGMDRPPFKLRDEF"  
ORIGIN  
Query Match 98.8%; Score 1086.8; DB 9; Length 1167;  
Best Local Similarity 99.4%; Pred. No. 1.5e-250;  
Matches 1091; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
DB 70 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 129  
QY 61 TTTGGGAACCAAGCCGATCCTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 120  
DB 130 TTTGGGAACCAAGCCGATCCTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 189  
QY 121 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCACAGCCCTCTTTACCAACCTC 180  
DB 190 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCACAGCCCTCTTTACCAACCTC 249  
QY 181 CATGTGTCCTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
DB 250 CATGTGTCCTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 309  
QY 241 AGCTTGGAGGATTTTCATGAGAGCTGCGACCCACCCACTGCCCTCTGAGAGCGGGTG 300  
DB 310 AGCTTGGAGGATTTTCATGAGAGCTGCGACCCACCCACTGCCCTCTGAGAGCGGGTG 369  
QY 301 GCATCTCTCTTGGAGTGGCAGCCGAGCCAGAGCCAGATTAAGAAGCTGCCCTCAAG 360  
DB 370 GCATCTCTCTTGGAGTGGCAGCCGAGCCAGAGCCAGATTAAGAAGCTGCCCTCAAG 429  
QY 361 GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
DB 430 GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 489  
QY 421 CTTTTCAGGCAATTTCTTCAAGTCTTCTTACAGAGAACATGAGCCAGAGATTTCT 480  
DB 490 CTTTTCAGGCAATTTCTTCAAGTCTTCTTACAGAGAACATGAGCCAGAGATTTCT 549  
QY 481 CCAAGGAACATCGGTCCTTCCCTGCGGAGCCGAGCCAGCCAGTTCCTCCCTCTAGAA 540  
DB 550 CCAAGGAACATCGGTCCTTCCCTGCGGAGCCGAGCCGAGCCAGTTCCTCCCTCTAGAA 609  
QY 541 GAACAAGCCCACTACAGAACTATGTTGTTGTCAGAGAAATGTTGAAGACGGAGAG 600  
DB 610 GAACAAGCCCACTACAGAACTATGTTGTTGTCAGAGAAATGTTGAAGACGGAGAG 669  
QY 601 GCCCAGATTCATGCCACCTTGTCCGCCCTTATGTGGGCTTCTGCGCATTCATGCGCT 660  
DB 670 GCCCAGATTCATGCCACCTTGTCCGCCCTTATGTGGGCTTCTGCGCATTCATGCGCT 729  
QY 661 GACTGGAAGACGCTGTGCCATGCTGAAGAGCGGAGCTGAGGCTCCGCTTCATGCGC 720  
DB 730 GACTGGAAGACGCTGTGCCATGCTGAAGAGCGGAGCTGAGGCTCCGCTTCATGCGC 789  
QY 721 TCTCCGAGTGTGTGGGCTACAGCCGAGCAGCGGCCCCCTCACGATGATATGTC 780  
DB 790 TCTCCGAGTGTGTGGGCTACAGCCGAGCAGCGGCCCCCTCACGATGATATGTC 849  
QY 781 CTGCTGACCTGAAGAGATCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
DB 850 CTGCTGACCTGAAGAGATCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 909  
QY 841 CAGTCGGTCTAGCTGCTACTGATTCCGAGAGTATGTGCGCTGAGCTCCAAAGCTCTTC 900

```
Db 910 CAGTCGGTCTATGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCACAGCTCTTC 969
Qy 901 AAAGGGAAGGTGAAGGTGTGAGCTGAAGCTGAGTGTGCGCCAGGTGACCTGTACATC 960
Db 970 AAAGGGAAGGTGAAGGTGTGAGCTGAAGCTGAGTGTGCGCCAGGTGACCTGTACATC 1029
Qy 961 CTCGGCCAAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTTGTGAAG 1020
Db 1030 CTCGGCCAAAGCTGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTTGTGAAG 1089
Qy 1021 CCGGAGCGGAGACCTCCAGGCGAGGCGCTCTCTCTTCTTCCGATGACAGGCGCCCTTAAG 1080
Db 1090 CCGGAGCGGAGACCTCCAGGCGAGGCGCTCTCTTCTTCCGATGACAGGCGCCCTTAAG 1149
Qy 1081 CTGCGGAGCAGGTCTGA 1098
Db 1150 CTGCGGAGCAGGTCTGA 1167

RESULT 14
AR163453
LOCUS AR163453 5009 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 8 from patent US 6270987.
ACCESSION AR163453
VERSION AR163453.1 GI:16234058
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5009)
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 8 07-AUG-2001;
FEATURES
    source
        Location/Qualifiers
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Query Match 93.9%; Score 1033.4; DB 6; Length 5009;
Best Local Similarity 99.9%; Pred. No. 1e-237;
Matches 1034; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 66 GAACGAGCCGATCAGCTTCTGGGCTCTGCGATTTGCAAGCTGCTAAACGCTACCTT 125
Db 1 GAACGAGCCGATCAGCTTCTGGGCTCTGCGATTTGCAAGCTGCTAAACGCTACCTT 60

Qy 126 GGTGTCTCTCTTGGATTGAGTACGAGATCACAGCTCTCTCTTTTCAACCACTCCATGT 185
Db 61 GGTGTCTCTCTTGGATTGAGTACGAGATCACAGCTCTCTCTTTTCAACCACTCCATGT 120

Qy 186 GTCTCTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATCAGCTT 245
Db 121 GTCTCTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATCAGCTT 180

Qy 246 GGAGGATTTTCATGAGAAGCTGGACCCACCTGCGCCCTGAGAGCGGTGGCATA 305
Db 181 GGAGGATTTTCATGAGAAGCTGGACCCACCTGCGCCCTGAGAGCGGTGGCATA 240

Qy 306 CTGCTTTGAGTGGGAGCCAGCAAGCCAGATGAAGAGCTGCCCATGAGGAAGG 365
Db 241 CTGCTTTGAGTGGGAGCCAGCAAGCCAGATGAAGAGCTGCCCATGAGGAAGG 300

Qy 366 AAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGTGATTTTCAACAGTCGGAGCTTTT 425
Db 301 AAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGTGATTTTCAACAGTCGGAGCTTTT 360

Qy 426 TACAGGCAATTTCTTCACTGCTTCTTACAGAGAACTGAGGAGCAAGATTTTCTCCAAA 485
Db 361 TACAGGCAATTTCTTCACTGCTTCTTACAGAGAACTGAGGAGCAAGATTTTCTCCAAA 420

Qy 486 GGAACATCGGTGCTTGCCTTGCAGGAGCCCGCCAGTTCCTCCGTCTCTAGAGAAC 545
```

```
Db 421 GGAACATCGGTGCTTGCCTTGCAGGAGCCCGCCAGTTCCTCCGTCTCTAGAGAAC 480
Qy 546 CAGGCCATCTACAGAGTACATGCTATGCTGATGCTGAGAGCAATGCTGAGAGGAGCCCA 605
Db 481 CAGGCCATCTACAGAGTACATGCTATGCTGATGCTGAGAGCAATGCTGAGAGGAGCCCA 540
Qy 606 GATTCAATGCCCACTTGTCTCGGCCCTTATGTGGGCAATTCATCTGCGCAATTTGGCTCTGACTG 665
Db 541 GATTCAATGCCCACTTGTCTCGGCCCTTATGTGGGCAATTCATCTGCGCAATTTGGCTCTGACTG 600
Qy 666 GAAGAACCCCTTGTGCCATGCTGAAGGACGAGGCTGCGAGGCTGCGACTTTTCATGGCTCTTCC 725
Db 601 GAAGAACCCCTTGTGCCATGCTGAAGGACGAGGCTGCGAGGCTGCGACTTTTCATGGCTCTTCC 660
Qy 726 GCAGTGTGTGGGCTTACAGCGCAGACAGCGGCCCTTCAAGATGACTATGTGCTGCG 785
Db 661 GCAGTGTGTGGGCTTACAGCGCAGACAGCGGCCCTTCAAGATGACTATGTGCTGCG 720
Qy 786 TGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCCAGTC 845
Db 721 TGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCCAGTC 780
Qy 846 GGTCTACCTTGTCTACTGATTCGAGAGTTTATGTGCTGAGCTCCACAGCTCTTCAAAGG 905
Db 781 GGTCTACCTTGTCTACTGATTCGAGAGTTTATGTGCTGAGCTCCACAGCTCTTCAAAGG 840
Qy 906 GAAGGTGAAGGTGTGAGCTGAAGCTGAGGTGCGCTGAGGTGCGTGGATGCCAGTC 965
Db 841 GAAGGTGAAGGTGTGAGCTGAAGCTGAGGTGCGCTGAGGTGCGTGGATGCCAGTC 900
Qy 966 CCAAGCGGACCACTTTATGGAAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1025
Db 901 CCAAGCGGACCACTTTATGGAAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Qy 1026 GCGGAGCTTCCAGGAGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1085
Db 961 GCGGAGCTTCCAGGAGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Qy 1086 GGACGAGTTCTGATT 1100
Db 1021 GGACGAGTTCTGATT 1035

RESULT 15
BD103191
LOCUS BD103191 5009 bp DNA linear PAT 27-AUG-2002
DEFINITION O-fucosyltransferase.
ACCESSION BD103191
VERSION BD103191.1 GI:22648765
KEYWORDS JP 2001527389-A/4.
SOURCE unidentified
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 5009)
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 4 25-DEC-2001;
COMMENT GENENTECH INC
OS Unidentified
PN JP 2001527389-A/4
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 JP 08/792498, 26-NOV-1997 US 08/978741 PI
YANG WANG, MICHAEL W SPELLMAN
PC C12N15/54, C12N9/10, C07K16/40
CC Strandedness: Single;
CC Topology: Linear;
CC O-fucosyltransferase
FT Key Location/Qualifiers
    source 1..5009 /organism='Unidentified'.
    FT Location/Qualifiers
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:23:13 ; Search time 642.158 Seconds  
(without alignments)  
10140.369 Million cell updates/sec

Title: US-09-774-954-16  
Perfect score: 1100  
Sequence: 1 ATGCCCGGGCTCTCTGGGA.....CTGCGGACGAGTTCGATT 1100

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description        |
|------------|--------|-------------|--------|----|--------------------|
| 1          | 1100   | 100.0       | 1514   | 2  | AAV65632 Human hea |
| 2          | 1098.4 | 99.9        | 11284  | 2  | AAV65633 Plasmid c |
| 3          | 1096.8 | 99.7        | 5218   | 13 | ACN40746 Tumour-as |
| 4          | 1095.8 | 99.6        | 5266   | 10 | ADF81754 Leukaemia |
| 5          | 1046.2 | 95.1        | 5230   | 4  | AAK51510 Human pol |
| 6          | 1041.4 | 94.7        | 4850   | 4  | AAK52494 Human pol |
| 7          | 1033.4 | 93.9        | 5009   | 2  | AAV65634 First Eco |
| 8          | 473.4  | 43.0        | 915    | 4  | AAV65634 Human sec |
| 9          | 406.2  | 36.9        | 610    | 13 | ADQ56824 Novel can |
| 10         | 251.2  | 22.8        | 10331  | 4  | AAK79915 Human imm |
| 11         | 251.2  | 22.8        | 10331  | 4  | AAK79915 Human imm |
| 12         | 233    | 21.2        | 479    | 9  | ACH32225 Human end |
| 13         | 221.8  | 20.2        | 477    | 9  | ACH35532 Human end |
| 14         | 166.4  | 15.1        | 1209   | 4  | ABL13947 Drosophil |
| c 15       | 150.4  | 13.7        | 3264   | 4  | ABL13946 Drosophil |
| c 16       | 123.2  | 11.2        | 1545   | 4  | AAF23894 Human sec |
| c 17       | 108.6  | 9.9         | 3793   | 4  | ABL13894 Drosophil |
| c 18       | 71.8   | 6.5         | 3567   | 4  | ABL13952 Drosophil |
| 19         | 60     | 5.5         | 60     | 6  | ABN44539 Human spl |
| 20         | 56     | 5.1         | 760    | 3  | AAZ98210 Human sig |

|      |      |     |        |    |          |                     |
|------|------|-----|--------|----|----------|---------------------|
| 21   | 50   | 4.5 | 2000   | 8  | ADA71938 | Ada71938 Rice gene  |
| c 22 | 46.2 | 4.2 | 2000   | 8  | ADA71938 | Ada71938 Rice gene  |
| 23   | 42.4 | 3.9 | 1438   | 8  | ACA31561 | ACA31561 Prokaryot  |
| c 24 | 41.2 | 3.7 | 114793 | 4  | AAD08215 | AAH97782 Murine gen |
| c 25 | 40.6 | 3.7 | 558    | 5  | AAH97782 | AAH97782 Murine gen |
| 26   | 40.4 | 3.7 | 5019   | 13 | ADQ80232 | Adq80232 Brain-spe  |
| 27   | 40.4 | 3.7 | 5704   | 12 | ADQ84111 | Adq84111 Human tum  |
| 28   | 40.4 | 3.7 | 5932   | 10 | ADB75488 | Adb75488 Prostate   |
| 29   | 40   | 3.6 | 40     | 2  | AAV65637 | AAV65637 Primer am  |
| 30   | 40   | 3.6 | 40     | 2  | AAV65635 | AAV65635 Northern   |
| 31   | 40   | 3.6 | 1554   | 8  | ADA69815 | ADA69815 Rice gene  |
| 32   | 39   | 3.5 | 372    | 8  | ACA43606 | ACA43606 Prokaryot  |
| 33   | 39   | 3.5 | 1131   | 11 | ABD01472 | ABD01472 Pseudomon  |
| 34   | 39   | 3.5 | 1353   | 11 | ABD01484 | ABD01484 Pseudomon  |
| 35   | 38.6 | 3.5 | 407    | 8  | ABX46567 | ABX46567 Bovine ES  |
| 36   | 38.6 | 3.5 | 1596   | 10 | ADI62782 | ADI62782 Human apo  |
| 37   | 38.6 | 3.5 | 2358   | 12 | ADE77025 | Ade77025 Human CDN  |
| 38   | 38.6 | 3.5 | 2670   | 10 | ADE25660 | Ade25660 Human adu  |
| 39   | 38.4 | 3.5 | 532    | 9  | ACH26305 | Ach26305 Human bre  |
| 40   | 38.4 | 3.5 | 6227   | 6  | ABT10150 | ABT10150 Human cor  |
| 41   | 38.4 | 3.5 | 6227   | 8  | ACA64963 | ACA64963 Human cor  |
| c 42 | 37.8 | 3.4 | 1234   | 13 | ADS63598 | ADS63598 Bacterial  |
| c 43 | 37.8 | 3.4 | 1234   | 13 | ADS63974 | ADS63974 Bacterial  |
| c 44 | 37.8 | 3.4 | 1234   | 13 | ADT41577 | Adt41577 Bacterial  |
| 45   | 37.8 | 3.4 | 2370   | 4  | AAH52044 | Aah52044 Mycobacte  |

ALIGNMENTS

RESULT 1  
AAV65632  
ID AAV65632 standard; DNA; 1514 BP.  
XX  
AC AAV65632;  
XX

16-DEC-1998 (first entry)  
XX Human heart O-fucosyltransferase encoding DNA.  
XX Human heart O-fucosyltransferase encoding DNA.

XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;  
KW O-fucose; inhibitor; sensory neuron; retinal neuron; heart; ss.  
XX Homo sapiens.

XX Key Location/Qualifiers  
FH misc\_feature 1..1100  
FT /tag= b  
FT /note= "this actively expressed O-fucosyltransferase  
sequence is claimed for in claim 9"  
FT 1..1098  
FT /tag= a  
FT /product= "human heart O-fucosyltransferase"

XX WO9833924-A1.  
XX 06-AUG-1998.  
XX 17-DEC-1997; 97WO-US023401.  
XX 31-JAN-1997; 97US-00792498.  
XX 26-NOV-1997; 97US-00978741.  
XX (GETH ) GENENTECH INC.

XX Wang Y, Spellman MW;  
XX WPI; 1998-437477/37.  
XX P-PSDB; AAW80571.

XX Human O-fucosyltransferase able to glycosylate epidermal growth factor  
domains - useful for diagnosis and treatment of diseases involving  
overexpression of the enzyme.

XX PS  
XX Claim 9; Fig 12A; 90pp; English.  
CC This DNA encodes a human heart O-fucosyltransferase that can glycosylate  
CC an epidermal growth factor (EGF) domain of a polypeptide with an  
CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.  
CC mutants with increased affinity for the EGF domain, are used in  
CC diagnosis and treatment of conditions associated with overexpression of O  
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.  
CC Probes based on EGF domain polypeptide are used to detect gene  
CC amplification and expression. The expression can also be determined at  
CC the protein level using antibodies specific for O-fucosyltransferase  
XX  
XX Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1100; DB 2; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 1.8e-306;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGCCCGCGGCTCTGGGACCGGCGGTTACTGCTCTACTGCCCCGTCATGGGGCGC 60  
Db 1 ATGCCCGCGGCTCTGGGACCGGCGGTTACTGCTCTACTGCCCCGTCATGGGGCGC 60  
Qy 61 TTTGGGAACACGAGCGCATCACCTTTCTGGGCTCTCTGGCATTTGCAAAAGCTGTAACCGT 120  
Db 61 TTTGGGAACACGAGCGCATCACCTTTCTGGGCTCTCTGGCATTTGCAAAAGCTGTAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCACAAAGCTCTCTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCACAAAGCTCTCTTACCAACCTC 180  
Qy 181 CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCTATC 240  
Db 181 CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCTATC 240  
Qy 241 AGCTTGGAGGATTTATGAGAGCTTGGACGCGACCCACTGGCCCCCTCAGAACGGGTG 300  
Db 241 AGCTTGGAGGATTTATGAGAGCTTGGACGCGACCCACTGGCCCCCTCAGAACGGGTG 300  
Qy 301 GCATATCTCTTTGAGTGGAGCCAGCGAAGCCAGATTAAGAAGACGTGCCCCATGAAG 360  
Db 301 GCATATCTCTTTGAGTGGAGCCAGCGAAGCCAGATTAAGAAGACGTGCCCCATGAAG 360  
Qy 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAAGGAACATCCGGTGTCTTGGCTGCGAGAGCCCGCCAGTTCCCGTCTCTAGAA 540  
Db 481 CCAAAGGAACATCCGGTGTCTTGGCTGCGAGAGCCCGCCAGTTCCCGTCTCTAGAA 540  
Qy 541 GAACACAGGCCACTACAGAGTACATGTTATGTTGTCAGACGAAATGGTGAAGACGGAGAG 600  
Db 541 GAACACAGGCCACTACAGAGTACATGTTATGTTGTCAGACGAAATGGTGAAGACGGAGAG 600  
Qy 601 GCCAGATTCATGCCCACTTGTGCGGCCCTATGTGGGCAATTCATCTGGGCAATGGCTCT 660  
Db 601 GCCAGATTCATGCCCACTTGTGCGGCCCTATGTGGGCAATTCATCTGGGCAATGGCTCT 660  
Qy 661 GACTGGAAGAACGCTGTGCTATGCTGAAGACGGGATCGAGGCTCGCACTTCATGCGC 720  
Db 661 GACTGGAAGAACGCTGTGCTATGCTGAAGACGGGATCGAGGCTCGCACTTCATGCGC 720  
Qy 721 TCTCCGAGTGTGGGCTACAGCGCGACAGCAGCGGCCCTCCCTCAGCATGACTATGTGC 780  
Db 721 TCTCCGAGTGTGGGCTACAGCGCGACAGCAGCGGCCCTCCCTCAGCATGACTATGTGC 780  
Qy 781 CTGGCTGACCTGAAGAGGATTCAGAGGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
|||||

Db 781 CTGGCTGACCTGAAGGAGATTCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGTTGATGCC 840  
Qy 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGATTATGTGCTCAGCTCCAAACAGCTCTTC 900  
Db 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGATTATGTGCTCAGCTCCAAACAGCTCTTC 900  
Qy 901 AAAGGGAAGGTGAAGGTGGTGGAGCTGAAGCTGAGGTGGCCAGGTCGACCTGTACATC 960  
Db 901 AAAGGGAAGGTGAAGGTGGTGGAGCTGAAGCTGAGGTGGCCAGGTCGACCTGTACATC 960  
Qy 961 CTCGGCCAAAGCCGACCACTTATTTGGCAACTGTGTCTCTCTCACTGCTTCTGTGAAG 1020  
Db 961 CTCGGCCAAAGCCGACCACTTATTTGGCAACTGTGTCTCTCTCTCACTGCTTCTGTGAAG 1020  
Qy 1021 CCGGACCGGAGCTTCAGGGGAGGCGCTTCTTCTTCTGGCATGGACAGGCCCTTAAG 1080  
Db 1021 CCGGACCGGAGCTTCAGGGGAGGCGCTTCTTCTTCTGGCATGGACAGGCCCTTAAG 1080  
Qy 1081 CTGGGACGAGTTCTGATT 1100  
Db 1081 CTGGGACGAGTTCTGATT 1100  
RESULT 2  
AAV65633  
ID AAV65633 standard; DNA; 11284 BP.  
AC AAV65633;  
XX  
XX  
XX 16-DEC-1998 (first entry)  
XX Plasmid construct for expression of human O-fucosyltransferase.  
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;  
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX  
XX Key Location/Qualifiers  
FT CDS 4140..5333  
FT /\*tag= a  
FT /note= "Insert coding for human O-fucosyltransferase."  
FT misc\_feature 4218..4235  
FT /\*tag= b  
FT /note= "polyhistidine tag"  
XX  
XX WO9833924-A1.  
XX  
XX 06-AUG-1998.  
XX  
XX 17-DEC-1997; 97WO-US023401.  
XX  
XX 31-JAN-1997; 97US-00792498.  
XX 26-NOV-1997; 97US-00978741.  
XX (GETH ) GENENTECH INC.  
XX  
XX Wang Y, Spellman MW;  
XX  
XX WPI; 1998-437477/37.  
XX P-PSDB; AAW80573.  
XX  
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor  
XX domains - useful for diagnosis and treatment of diseases involving  
XX overexpression of the enzyme.  
XX  
XX Example; Page 49-57; 90pp; English.  
XX  
XX This represents the nucleotide sequence of the plasmid construct used for  
XX the expression of human heart O-fucosyltransferase. The human O-  
XX fucosyltransferase can glycosylate an epidermal growth factor (EGF)  
XX domain of a polypeptide with an activated O-fucose residue. Inhibitors of

CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG  
CC domains, are used in diagnosis and treatment of conditions associated  
CC with overexpression of O-fucosyltransferase, to promote survival of  
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are  
CC used to detect gene amplification and expression. The expression can also  
CC be determined at the protein level using antibodies specific for O-  
CC fucosyltransferase  
XX

SQ Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;

Query Match 99.9%; Score 1098.4; DB 2; Length 11284;  
Best Local Similarity 99.9%; Pred. No. 1.3e-305;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

|    |      |  |      |
|----|------|--|------|
| Qy | 1    | ATGCCCGCGGGCTCTCTGGGACCCGCGGTACCTGCTCTACTGCGCCCTGATGGGGCGC   | 60   |
| Db | 4236 | ATGCCCGCGGGCTCTCTGGGACCCGCGGTACCTGCTCTACTGCGCCCTGATGGGGCGC   | 4295 |
| Qy | 61   | TTTGGGAACCAAGCGCATCACTTCTTGGGCTCTCTGGCAITTTGCAAGCTGCTAAACCGT | 120  |
| Db | 4296 | TTTGGGAACCAAGCGCATCACTTCTTGGGCTCTCTGGCAITTTGCAAGCTGCTAAACCGT | 4355 |
| Qy | 121  | ACCTTGGCTGCTCCCTCTTGATGAGTACAGCATCAAGACCTCTTTTACCAACCTC      | 180  |
| Db | 4356 | ACCTTGGCTGCTCCCTCTTGATGAGTACAGCATCAAGACCTCTTTTACCAACCTC      | 4415 |
| Qy | 181  | CATGTGCTCTACCAAGTACTTCAAGCTGAGGCCCTCCAGGCTTACCATCGGTCATC     | 240  |
| Db | 4416 | CATGTGCTCTACCAAGTACTTCAAGCTGAGGCCCTCCAGGCTTACCATCGGTCATC     | 4475 |
| Qy | 241  | AGCTTGGAGGATTTTCATGGAAGCTGCGACCCACCCACTGCGCCCTTGAGAAGCGGGTG  | 300  |
| Db | 4476 | AGCTTGGAGGATTTTCATGGAAGCTGCGACCCACCCACTGCGCCCTTGAGAAGCGGGTG  | 4535 |
| Qy | 301  | GCATCTGCTTTGAGTGGAGCCAGCAAGCCAGATAGAGAGAGCGTCCCATGAG         | 360  |
| Db | 4536 | GCATCTGCTTTGAGTGGAGCCAGCAAGCCAGATAGAGAGAGCGTCCCATGAG         | 4595 |
| Qy | 361  | GAAGAAACCCCTTTGGCCCACTTCGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG   | 420  |
| Db | 4596 | GAAGAAACCCCTTTGGCCCACTTCGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG   | 4655 |
| Qy | 421  | CTTTTTCAGGCAITTCCTTCAGTGTCTTCTACAGAGAAATGAGCGCAGAGATTTTCT    | 480  |
| Db | 4656 | CTTTTTCAGGCAITTCCTTCAGTGTCTTCTACAGAGAAATGAGCGCAGAGATTTTCT    | 4715 |
| Qy | 481  | CCAAAGGAACATCGGTGCTTCCCTGCGGAGCCCGCCAGCTTCCCGCTCTAGAA        | 540  |
| Db | 4716 | CCAAAGGAACATCGGTGCTTCCCTGCGGAGCCCGCCAGCTTCCCGCTCTAGAG        | 4775 |
| Qy | 541  | GAACACAGGCCCACTACAGAGTACATGTATGTTGTCAGACGAATGGTGAACGGGAGAG   | 600  |
| Db | 4776 | GAACACAGGCCCACTACAGAGTACATGTATGTTGTCAGACGAATGGTGAACGGGAGAG   | 4835 |
| Qy | 601  | GCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGCAITTCATCTGCGCAITTTGCTCT | 660  |
| Db | 4836 | GCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGCAITTCATCTGCGCAITTTGCTCT | 4895 |
| Qy | 661  | GACTGGAAGACCGCTGTGCCATGCTGAAGAGCGGACTGCGCTGCGCATTCATGGCC     | 720  |
| Db | 4896 | GACTGGAAGACCGCTGTGCCATGCTGAAGAGCGGACTGCGCTGCGCATTCATGGCC     | 4955 |
| Qy | 721  | TCTCCGAGTGTGGGTACAGCGGAGCAGACGCGCCCTCCAGATGACTATGTGC         | 780  |
| Db | 4956 | TCTCCGAGTGTGGGTACAGCGGAGCAGACGCGCCCTCCAGATGACTATGTGC         | 5015 |
| Qy | 781  | CTGCTGACCTGGAAGGAGATCCAGAGGCTCTGGAAGCTCTGGGTGAGTCCCTGGATGCC  | 840  |
| Db | 5016 | CTGCTGACCTGGAAGGAGATCCAGAGGCTCTGGAAGCTCTGGGTGAGTCCCTGGATGCC  | 5075 |
| Qy | 841  | CAGTCGGTCTACGTGCTTCTGATTCGAGAGGTTATGTGCTGTAGCTCCACAGCTCTTC   | 900  |
| Db | 5076 | CAGTCGGTCTACGTGCTTCTGATTCGAGAGGTTATGTGCTGTAGCTCCACAGCTCTTC   | 5135 |

## RESULT 3

|          |    |  |                                 |
|----------|----|--|---------------------------------|
| ACN40746 | ID | ACN40746   | standard; cDNA; 5218 BP.        |
| XX       | AC | ACN40746;  |                                 |
| XX       | DT | 18-NOV-2004  | (first entry)                   |
| XX       | XX | Tumour-associated antigenic target (TAT)                                 | CDNA DNA326813, SEQ ID NO:5725. |
| XX       | DE | Tumour-associated antigenic target; TAT; human; overexpression; cancer;  |                                 |
| KW       | KW | tumour; diagnosis; cell proliferative disorder; breast cancer;           |                                 |
| KW       | KW | colorectal cancer; lung cancer; ovarian cancer; liver cancer;            |                                 |
| KW       | KW | central nervous system cancer; bladder cancer; pancreatic cancer;        |                                 |
| KW       | KW | cervical cancer; melanoma; leukaemia; hybridisation probe;               |                                 |
| KW       | KW | chromosome identification; chromosome mapping; gene mapping;             |                                 |
| KW       | KW | gene therapy; cytostatic; gene; ss.                                      |                                 |
| XX       | OS | Homo sapiens.  |                                 |
| XX       | PN | WO2004030615-A2.   |                                 |
| XX       | XX | 15-APR-2004.   |                                 |
| XX       | PD | 29-SEP-2003; 2003WO-US028547.  |                                 |
| XX       | PF | 02-OCT-2002; 2002US-0414971P.  |                                 |
| XX       | PR | (GETH ) GENENTECH INC.   |                                 |
| XX       | PA | Wu TD, Zhang Z, Zhou Y;  |                                 |
| XX       | PI | WPI; 2004-347921/32.   |                                 |
| XX       | PI | P-PSDB; ABM82223.  |                                 |
| XX       | DR | New tumor-associated antigenic target polypeptides and nucleic acids,    |                                 |
| XX       | DR | useful in preparing a medicament for treating or detecting a             |                                 |
| XX       | DR | proliferative disorder, e.g. breast, lung, colorectal, ovarian or        |                                 |
| XX       | DR | prostate cancer or tumor.  |                                 |
| XX       | PT | Claim 1; SEQ ID NO 5725; 7273pp; English.                                |                                 |
| XX       | PS | The invention relates to human tumour-associated antigenic target (TAT)  |                                 |
| XX       | XX | polypeptides, and their related nucleic acids. The TAT polypeptides are  |                                 |
| CC       | CC | overexpressed in cancer tissues compared to normal tissues, and may thus |                                 |
| CC       | CC | serve as effective targets for the diagnosis and treatment of cancer in  |                                 |
| CC       | CC | mammals. The invention also relates to nucleic acid and polypeptide      |                                 |
| CC       | CC | sequences at least 80% identical to the TAT nucleic acids and            |                                 |
| CC       | CC | polypeptides; expression vectors and host cells comprising a TAT nucleic |                                 |
| CC       | CC | acid; an antibody specific for a TAT polypeptide; a peptide or organic   |                                 |
| CC       | CC | molecule which binds to a TAT polypeptide; fusion proteins comprising a  |                                 |
| CC       | CC | TAT polypeptide; and methods and compositions for the treatment or       |                                 |
| CC       | CC | diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,         |                                 |
| CC       | CC | antibodies, antagonists, binding molecules and compositions are useful   |                                 |
| CC       | CC | for diagnosing or treating a cell proliferative disorder associated with |                                 |

CC increased TAT expression, particularly cancers such as breast cancer, colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder cancer, pancreatic cancer, cervical cancer, cancers of the central nervous system, melanoma and leukaemia. TAT nucleic acids may further be used as hybridisation probes, in chromosome and gene mapping, in chromosome identification and in gene therapy. The present sequence represents a TAT nucleic acid of the invention

XX  
SQ Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;

Query Match 99.7%; Score 1096.8; DB 13; Length 5218;  
Best Local Similarity 99.8%; Pred. No. 2.7e-305;  
Matches 1098; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCCCGGGCTCTGGAGACCGCGGTACTCGCTCTACTGCCCTGCATGGGGCGC 60  
Db |||  
Qy 61 TTGTGGAAACAGGCGCGATCATCTTTGGGCTCTCTGGCATTTGCCAAGCTGTAACCGT 120  
Db |||  
Qy 179 TTGTGGAAACAGGCGCGATCATCTTTGGGCTCTCTGGCATTTGCCAAGCTGTAACCGT 238  
Db |||  
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACCAGCATCACAAAGCTCTCTTTCACCAACCTC 180  
Db |||  
Qy 239 ACCTTGGCTGTCCCTCTCTGGATTGAGTACCAGCATCACAAAGCTCTCTTTCACCAACCTC 298  
Db |||  
Qy 181 CATGTGTCTTACCAAGAGTACTTCAAGCTTGGAGCCCTTCCAGGCTTACCATCGGTCTATC 240  
Db |||  
Qy 299 CATGTGTCTTACCAAGAGTACTTCAAGCTTGGAGCCCTTCCAGGCTTACCATCGGTCTATC 358  
Db |||  
Qy 241 ACCTTGGAGGATTTTCATGAGAGCTTGGCACCACCTGCGCCCTCAGAAAGCGGGTG 300  
Db |||  
Qy 359 ACCTTGGAGGATTTTCATGAGAGCTTGGCACCACCTGCGCCCTCAGAAAGCGGGTG 418  
Db |||  
Qy 301 GCATCTCTCTTGGAGTGGCGCCAGAGAGCCAGATGAAGAGCTGCCCCATGAAG 360  
Db |||  
Qy 419 GCATCTCTCTTGGAGTGGCGCCAGAGAGCCAGATGAAGAGCTGCCCCATGAAG 478  
Db |||  
Qy 361 GAAGAAACCCCTTTGGCCCATCTGGGATCATGTTTATGTGAGTTTCAAGTCGGAG 420  
Db |||  
Qy 479 GAAGAAACCCCTTTGGCCCATCTGGGATCATGTTTATGTGAGTTTCAAGTCGGAG 538  
Db |||  
Qy 421 CTTTTTACAGCATTTCTTCACTAGTGTCTCTACAGAGAAATGAGAGCATATTTCT 480  
Db |||  
Qy 539 CTTTTTACAGCATTTCTTCACTAGTGTCTCTACAGAGAAATGAGAGCATATTTCT 598  
Db |||  
Qy 481 CCAAAGGAACATCCGGTGTCTTGGCCCTGCGAGAGCCCGCCAGTTCCCGCTCTAGAA 540  
Db |||  
Qy 599 CCAAAGGAACATCCGGTGTCTTGGCCCTGCGAGAGCCCGCCAGTTCCCGCTCTAGAG 658  
Db |||  
Qy 541 GAACACAGCCACTACAGAAATACATGTATGTGTGAGAGAAATGAGAGCGGGAGAG 600  
Db |||  
Qy 659 GAACACAGCCACTACAGAAATACATGTATGTGTGAGAGAAATGAGAGCGGGAGAG 718  
Db |||  
Qy 601 GCCAGATTTCATGCCACCTTGTCCGCGCCCTATGTGGGCAATTCATCTGCGCATTTGCTCT 660  
Db |||  
Qy 719 GCCAGATTTCATGCCACCTTGTCCGCGCCCTATGTGGGCAATTCATCTGCGCATTTGCTCT 778  
Db |||  
Qy 661 GACTGAGAGAGCGCTGTGTCATGCTGAGAGAGCGGACTGCGACTGCGACTTCATGCC 720  
Db |||  
Qy 779 GACTGAGAGAGCGCTGTGTCATGCTGAGAGAGCGGACTGCGACTGCGACTTCATGCC 838  
Db |||  
Qy 721 TCTCCGAGTGTGGGCTACAGCCGAGACACAGCGGCCCTCTCAGATGACTATGTGC 780  
Db |||  
Qy 839 TCTCCGAGTGTGGGCTACAGCCGAGACACAGCGGCCCTCTCAGATGACTATGTGC 898  
Db |||  
Qy 781 CTGCTGTACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGTGCTGTGATGCC 840  
Db |||  
Qy 899 CTGCTGTACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGTGCTGTGATGCC 958  
Db |||  
Qy 841 CAGTCGGTCTAGTGTCTTACTGATTCGAGAGTATATGTGCTGTGAGTCTCAACAGCTCTTC 900  
Db |||  
Qy 959 CAGTCGGTCTAGTGTCTTACTGATTCGAGAGTATATGTGCTGTGAGTCTCAACAGCTCTTC 1018  
Db |||

Qy 901 AAAGGAGAGTGAAGTGTGAGCTTGAAGCTGAGTGCCGAGGTGAGCTGACCTGACCTGATC 960  
Db |||  
Qy 1019 AAAGGAGAGTGAAGTGTGAGCTTGAAGCTGAGTGCCGAGGTGAGCTGACCTGATC 1078  
Db |||  
Qy 961 CTCGGCCAAAGCCAGCACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAG 1020  
Db |||  
Qy 1079 CTCGGCCAAAGCCAGCACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAG 1138  
Db |||  
Qy 1021 CGGAGAGCGGAGCTTCCAGGGAGGCGCTTCTTCTTCTTTCGGCATGACAGGCCCCCTAAG 1080  
Db |||  
Qy 1139 CGGAGAGCGGAGCTTCCAGGGAGGCGCTTCTTCTTCTTTCGGCATGACAGGCCCCCTAAG 1198  
Db |||  
Qy 1081 CTGCGGGAGAGTTCGTGATT 1100  
Db |||  
Qy 1199 CTGCGGGAGAGTTCGTGATT 1218  
Db |||

RESULT 4  
ADF81754  
ID ADF81754 standard; DNA; 5266 BP.  
XX  
AC ADF81754;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Leukaemia-Related DNA sequence #2310.  
XX  
KW Cytostatic; Gene therapy; leukaemia; ss.  
XX  
OS Unidentified.  
XX  
PN W02003039443-A2.  
XX  
PD 15-MAY-2003.  
XX  
PF 04-NOV-2002; 2002WO-EP012303.  
XX  
PR 05-NOV-2001; 2001EP-00126244.  
XX  
PR 30-APR-2002; 2002EP-00009758.  
XX  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
PA (UFLU-) UNIV LUDWIG MAXIMILIANS.  
PA (HAFE/) HAFERLACH T.  
PA (SCHO/) SCHOCH C.  
PA (KERN/) KERN W.  
XX  
XX Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;  
PI Eils R, Brors B, Mergenthaler S;  
XX  
DR WPI; 2003-505037/47.  
XX  
PT Determining the subtype of leukemia cells and whether a patient sample  
PT contains leukemia cells or other cells, useful for treating leukemia,  
PT comprises determining the expression profile of a group of markers in a  
PT patient sample.  
XX  
PS Disclosure; SEQ ID NO 2310; 2938pp; English.  
XX  
CC The present invention relates to a method (M1) for determining the  
CC subtype of leukaemia cells and whether a patient sample contains  
CC leukaemia cells. The method comprises determining the expression profile  
CC of a group of markers in a patient sample. The method is useful for  
CC determining the presence of leukaemia cells, its type or subtypes, and  
CC for the preparation of a medicament for treating leukaemia.  
XX  
SQ Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;

Query Match 99.6%; Score 1095.8; DB 10; Length 5266;  
Best Local Similarity 99.7%; Pred. No. 5.2e-305;  
Matches 1097; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGCCCGGGCTCTGGAGACCGCGGTACTCGCTCTACTGCCCTGCATGGGGCGC 60



Dd 1212 CTGCGGGACAGATTCTTGATT 1231

RESULT 5  
AAK51510  
ID AAK51510 standard; cDNA; 5230 BP.  
XX AC AAK51510;  
XX XX  
XX 06-NOV-2001 (first entry)  
XX XX  
DE Human polynucleotide SEQ ID NO 55.  
XX XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX OS Homo sapiens.  
XX XX  
XX W0200157190-A2.  
XX XX  
XX 09-AUG-2001.  
XX PD  
XX PF 05-FEB-2001; 2001WO-US004098.  
XX PR 03-FEB-2000; 2000US-00496914.  
XX PR 27-APR-2000; 2000US-00560875.  
XX PR 20-JUN-2000; 2000US-00598075.  
XX PR 19-JUL-2000; 2000US-00620325.  
XX PR 01-SEP-2000; 2000US-00654936.  
XX PR 15-SEP-2000; 2000US-00663561.  
XX PR 20-OCT-2000; 2000US-00693325.  
XX PR 30-NOV-2000; 2000US-00728422.  
XX XX  
FA (HYSE-) HYSEQ INC.  
XX XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX XX  
DR WPI; 2001-476283/51.  
XX DR P-PSDB; AAM78377.  
XX XX  
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
XX XX  
FS Claim 1; Page 638-642; 6221pp; English.  
XX XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX XX  
SQ Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;  
XX XX  
Query Match 95.1%; Score 1046.2; DB 4; Length 5230;  
Best Local Similarity 99.7%; Pred. No. 1.1e-290;  
Matches 1048; Conservative 0; Mismatches 3; Indels 0; Gaps 0  
  
Qy 50 GCATGGGGCGCTTTGGGAACAGGCCGATCATCTTCTTGGGTCTCTGCATTGGCAAGC 109  
|||  
Dd 206 GCCAGGGGCGCTTTGGGAACAGGCCGATCATCTTCTTGGGTCTCTGCATTGGCAAGC 265

QY 110 TGCTAAACCGTACTTGGCTGCTCCCTCTTGATGATGATACAGCATCAAGCTCCTTT 169  
Db 266 TGCTAAACCGTACTTGGCTGCTCCCTCTTGATGATGATACAGCATCAAGCTCCTTT 325  
QY 170 TCACCAACCTCCATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACC 229  
Db 326 TCACCAACCTCCATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACC 385  
QY 230 ATCCGGTTCATCAGCTTGGAGGATTTTCATGAGAGCTGGCACCACCACTGGCCCTG 289  
Db 386 ATCCGGTTCATCAGCTTGGAGGATTTTCATGAGAGCTGGCACCACCACTGGCCCTG 445  
QY 290 AGAAGCGGTGGCATACTGCTTTGAGGTGGCAGCCAGCCAGCCAGATGAGAGCGT 349  
Db 446 AGAAGCGGTGGCATACTGCTTTGAGGTGGCAGCCAGCCAGCCAGATGAGAGCGT 505  
QY 350 GCCCATGAAGAGGAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCA 409  
Db 506 GCCCATGAAGAGGAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCA 565  
QY 410 ACAAGTCGAGCTTTTTCAGAGCATTTCTTTCAGTGTCTTCTACAGAGAACTGGAGCC 469  
Db 566 ACNAGTCGAGCTTTTTCAGAGCATTTCTTTCAGTGTCTTCTACAGAGAACTGGAGCC 625  
QY 470 AGAGATTTTCTCAAAGGAACATCGGTGCTTGGCCCTGAGAGCCAGCCAGCTTCC 529  
Db 626 AGAGATTTTCTCAAAGGAACATCGGTGCTTGGCCCTGAGAGCCAGCCAGCTTCC 685  
QY 530 CGCTCTAGAGAACACAGGCCATCTACAGAGTATCGTATGGTCAGACGAAATGGTGA 589  
Db 686 CGCTCTAGAGAACACAGGCCATCTACAGAGTATCGTATGGTCAGACGAAATGGTGA 745  
QY 590 AGACGGGAGAGCCAGATTCATGCCACCTTTGTCGGCCCTATGTGGGCAATTCATCTGC 649  
Db 746 AGACGGGAGAGCCAGATTCATGCCACCTTTGTCGGCCCTATGTGGGCAATTCATCTGC 805  
QY 650 GCATTGGCTCTGACTGGAAGACGCTTGCCATCTGAGGACGAGCTGCAGGCTGCG 709  
Db 806 GCATTGGCTCTGACTGGAAGACGCTTGCCATCTGAGGACGAGCTGCAGGCTGCG 865  
QY 710 ACTTCATGGCTCTGCGCAGTGTGGGCTACAGCCGAGCAGAGCCGCCCTTCAGGA 769  
Db 866 ACTTCATGGCTCTGCGCAGTGTGGGCTACAGCCGAGCAGAGCCGCCCTTCAGGA 925  
QY 770 TGACTATGTGCTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 829  
Db 926 TGACTATGTGCTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 985  
QY 830 CGCTGGATGCCAGTGGCTTACGTTGCTTACTGATTCGAGAGTTATGTGCTGAGCTCC 889  
Db 986 CGCTGGATGCCAGTGGCTTACGTTGCTTACTGATTCGAGAGTTATGTGCTGAGCTCC 1045  
QY 890 AACAGCTTTCAAAGGAGAGTGAAGTGTGAGCTGAGCCCTGAGCTGAGCTGGCCAGTTCG 949  
Db 1046 AACAGCTTTCAAAGGAGAGTGAAGTGTGAGCTGAGCCCTGAGCTGAGCTGGCCAGTTCG 1105  
QY 950 ACTGTATCATCTCGGCCAAGCCAGCCACTTTATTTGGCACTGTGTCTCTCTTCACTG 1009  
Db 1106 ACTGTATCATCTCGGCCAAGCCAGCCACTTTATTTGGCACTGTGTCTCTCTTCACTG 1165  
QY 1010 CTTTGTGAAGCGGAGCGGAGCCTCCAGGGAGAGCCGTCTTCTTCTTTCGGCATGGACA 1069  
Db 1166 CTTTGTGAAGCGGAGCGGAGCCTCCAGGGAGAGCCGTCTTCTTCTTTCGGCATGGACA 1225  
QY 1070 GGGCCCTTAAGCTGGGAGAGTCTGATT 1100  
Db 1226 GGGCCCTTAAGCTGGGAGAGTCTGATT 1256

RESULT 6  
AAK52494  
ID AAK52494 standard; cDNA; 4850 BP.  
XX

AAK52494;  
06-NOV-2001 (first entry)  
Human polynucleotide SEQ ID NO 2023.  
Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
tissue growth factor; immunomodulatory; cancer; leukaemia;  
nervous system disorder; arthritis; inflammation; ss.  
Homo sapiens.  
W0200157190-A2.  
09-AUG-2001.  
05-FEB-2001; 2001WO-US004098.  
03-FEB-2000; 2000US-00496914.  
27-APR-2000; 2000US-00560875.  
20-JUN-2000; 2000US-00598075.  
19-JUL-2000; 2000US-00620325.  
01-SEP-2000; 2000US-00654936.  
15-SEP-2000; 2000US-00663561.  
20-OCT-2000; 2000US-00693325.  
30-NOV-2000; 2000US-00728422.  
(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
WPI; 2001-476283/51.  
P-PSDB; AAM79361.  
Nucleic acids encoding polypeptides with cytokine-like activities, useful  
in diagnosis and gene therapy.  
Claim 1; Page 4424-4425; 6221pp; English.  
The invention relates to polynucleotides (AAK51456-AAK53435) and the  
encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
cytokine, cell proliferation or cell differentiation or which may induce  
production of other cytokines in other cell populations. The  
polynucleotides and polypeptides are useful in gene therapy, vaccines or  
peptide therapy. The polypeptides have various cytokine-like activities,  
e.g. stem cell growth factor activity, haematopoiesis regulating  
activity, tissue growth factor activity, immunomodulatory activity and  
activin/inhibin activity and may be useful in the diagnosis and/or  
treatment of cancer, leukaemia, nervous system disorders, arthritis and  
inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
(AAK52592) and 3666 (AAM80020) are omitted as the relevant pages from the  
sequence listing were missing at the time of publication  
Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;  
Query Match 94.7%; Score 1041.4; DB 4; Length 4850;  
Best Local Similarity 99.4%; Pred. No. 2.5e-289;  
Matches 1045; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 50 GCATGGGGCGCTTTGGGAAACAGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGC 109  
Db 206 GCCAGGGGCGCTTTGGGAAACAGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGC 265  
QY 110 TGTAAACCGTACCTTGGCTGCTCCCTTGGATGAGTACAGCATCAAGCTCCTTT 169  
Db 266 TGTGNAACCGTTCCTTGGCTGCTCCCTTGGATGAGTACAGCATCAAGCTCCTTT 325  
QY 170 TCACCAACCTCCATGTGTCTTACAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACC 229  
Db 326 TCACCAACCTCCATGTGTCTTACAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACC 385

|    |      |   |      |
|----|------|---|------|
| QY | 230  | ATCGGGTCATCAGCTTGGAGGATTTTCATGAGAAAGCTGGCAACCACCTCGGCCCTCG    | 289  |
| DB | 386  | ATCGGGTCATCAGCTTGGAGGATTTTCATGAGAAAGCTGGCAACCACCTCGGCCCTCG    | 445  |
| QY | 290  | AGAAAGCGGTGGCATACTGCTTTTGAGTGGAGCCACGAGAGCCACAGTAAGAAGAGCT    | 349  |
| DB | 446  | AGAAAGCGGTGGCATACTGCTTTTGAGTGGAGCCACGAGAGCCACAGTAAGAAGAGCT    | 505  |
| QY | 350  | GCCCCATGAAGAAAGGAAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGATTTCA    | 409  |
| DB | 506  | GCCCCATGAAGAAAGGAAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGATTTCA    | 565  |
| QY | 410  | ACAAGTCGGAGCTTTTACAGGCATTTCCCTTCAGTGCTTCTTACAGAGAACAATGGAGCC  | 469  |
| DB | 566  | ACAAGTCGGAGCTTTTACAGGCATTTCCCTTCAGTGCTTCTTACAGAGAACAATGGAGCC  | 625  |
| QY | 470  | AGAGATTTTCTCAAAGGAAACATCCGGTGCTTGCCTGCCAGAGAGCCCCAGCCAGTTCC   | 529  |
| DB | 626  | AGAGATTTTCTCAAAGGAAACATCCGGTGCTTGCCTGCCAGAGAGCCCCAGCCAGTTCC   | 685  |
| QY | 530  | CCGCTCTAGAAACACACAGGCCACTACAGAAAGTACATGGTATGGTCAGACGAAATGGTGA | 589  |
| DB | 686  | CCGCTCTAGAGAAACACAGGCCACTACAGAAAGTACATGGTATGGTCAGACGAAATGGTGA | 745  |
| QY | 590  | AGACGGGAGAGCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGC    | 649  |
| DB | 746  | AGACGGGAGAGCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGC    | 805  |
| QY | 650  | GCAATTGGCTCTGACTGGAAAGAACGCTGTGCCATGCTGAAGACGGGACTCGAGGCTCGC  | 709  |
| DB | 806  | GCAATTGGCTCTGACTGGAAAGAACGCTGTGCCATGCTGAAGACGGGACTCGAGGCTCGC  | 865  |
| QY | 710  | ACTTCATGGCCTCTCCGAGTGTGTGGGCTACAGCCGACACACGCGCCCCCTCACGA      | 769  |
| DB | 866  | ACTTCATGGCCTCTCCGAGTGTGTGGGCTACAGCCGACACACGCGCCCCCTCACGA      | 925  |
| QY | 770  | TGACTATGTGCTCGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGT     | 829  |
| DB | 926  | TGACTATGTGCTCGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGT     | 985  |
| QY | 830  | CGCTGGATGCCAGTCCGCTTACGTTTGTCTGATTCGAGAGTTATGTGCTCGAGCTCC     | 889  |
| DB | 986  | CGCTGGATGCCAGTCCGCTTACGTTTGTCTGATTCGAGAGTTATGTGCTCGAGCTCC     | 1045 |
| QY | 890  | AACAGCTCTTCAAAGGGAAGGTGAAGTGGTGAGCCTGAAGCCTGAGTGGCCCCAGGTGC   | 949  |
| DB | 1046 | AACAGCTCTTCAAAGGGAAGGTGAAGTGGTGAGCCTGAAGCCTGAGTGGCCCCAGGTGC   | 1105 |
| QY | 950  | ACCTGTACATCCTCGGCCAAGCCCAACATTTTATTTGGCAACTGTGTCTCTCTTCACTG   | 1009 |
| DB | 1106 | ACCTGTACATCCTCGGCCAAGCCCAACATTTTATTTGGCAACTGTGTCTCTCTTCACTG   | 1165 |
| QY | 1010 | CCTTTGTGAAGCGGAGCGGAGCCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA   | 1069 |
| DB | 1166 | CCTTTGTGAAGCGGAGCGGAGCCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA   | 1225 |
| QY | 1070 | GGCCCCCTTAAGCTCGGGAAGAGTTCTGATT                               | 1100 |
| DB | 1226 | GGCCCCCTTAAGCTCGGGAAGAGTTCTGATT                               | 1256 |

|          |  |
|----------|--|
| RESULT 7 |  |
| AAV65634 |  |
| ID       | AAV65634 standard; cDNA; 5009 BP.                                  |
| XX       |  |
| XX       | AAV65634;  |
| XX       |  |
| XX       |  |
| DT       | 16-DEC-1998 (first entry)  |
| XX       |  |
| XX       |  |
| DE       | First EcoRI nucleotide fragment of human KIAA0180.                 |
| XX       |  |
| XX       |  |
| KW       | O-fucosyltransferase; epidermal growth factor; EGF; glycosylation; |

KW O-fucose; inhibitor; sensory neuron; retinal neuron; heart; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9833924-A1.  
 XX  
 PD 06-AUG-1998.  
 XX  
 XX  
 PF 17-DEC-1997; 97WO-US023401.  
 XX  
 XX 31-JAN-1997; 97US-00792498.  
 PR 26-NOV-1997; 97US-00978741.  
 PR  
 XX (GETH ) GENENTECH INC.  
 PA  
 XX Wang Y, Spellman MW;  
 PI  
 XX WPI; 1998-437477/37.  
 DR  
 DR P-PSDB; AAW80577.  
 XX  
 XX  
 PT Human O-fucosyltransferase able to glycosylate epidermal growth factor  
 PT domains - useful for diagnosis and treatment of diseases involving  
 PT overexpression of the enzyme.  
 PS  
 PS Example 1; Fig 11; 90pp; English.  
 XX  
 CC This represents a first EcoRI nucleotide fragment of human KIAA0180. This  
 CC 5009 basepairs partial cDNA encodes for a protein of unknown function  
 CC from myeloblast cell line KG-1. The invention provides a human heart O-  
 CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)  
 CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of  
 CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF  
 CC domains, are used in diagnosis and treatment of conditions associated  
 CC with overexpression of O-fucosyltransferase, to promote survival of  
 CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are  
 CC used to detect gene amplification and expression. The expression can also  
 CC be determined at the protein level using antibodies specific for O-  
 CC fucosyltransferase  
 XX  
 SQ Sequence 5009 BP: 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

| Seq                   | Sequence     | 5009 BP;                    | 1235 A;                     | 1195 C;             | 1268 G; | 1311 T; | 0 U;  | 0 Other; |
|-----------------------|--------------|-----------------------------|-----------------------------|---------------------|---------|---------|-------|----------|
| Query Match           | 93.9%;       | Score                       | 1033.4;                     | DB                  | 2;      | Length  | 5009; |          |
| Best Local Similarity | 99.9%;       | Pred. No.                   | 5.1e-287;                   |                     |         |         |       |          |
| Matches 1034;         | Conservative | 0;                          | Mismatches                  | 1;                  | Indels  | 0;      | Gaps  | 0;       |
| Qy                    | 66           | GAACCAAGCCGATC              | ACTTCTTGGGCTCTCTGGGCATTTGCA | AGCTGCTAAACCGTACCTT | 125     |         |       |          |
| Db                    | 1            | GAACCAAGCCGATC              | ACTTCTTGGGCTCTCTGGGCATTTGCA | AGCTGCTAAACCGTACCTT | 60      |         |       |          |
| Qy                    | 126          | GGCTGTCCCTCCTTGGATTGAGTACCA | GAGCATCACAAGCCTCCTTTCA      | CAACCTCCATGT        | 185     |         |       |          |
| Db                    | 61           | GGCTGTCCCTCCTTGGATTGAGTACCA | GAGCATCACAAGCCTCCTTTCA      | CAACCTCCATGT        | 120     |         |       |          |
| Qy                    | 186          | GTCTTACCAAGATGACTTCAAGCTGG  | AGCCCTCCTCAGGTTTACATCGG     | GTATCAGTT           | 245     |         |       |          |
| Db                    | 121          | GTCTTACCAAGATGACTTCAAGCTGG  | AGCCCTCCTCAGGTTTACATCGG     | GTATCAGTT           | 180     |         |       |          |
| Qy                    | 246          | GGAGGATTTTCATGGAGAAAGCTGG   | CACCCACCCACTGGCCCCCTCAGA    | AGCGGGTGGCATA       | 305     |         |       |          |
| Db                    | 181          | GGAGGATTTTCATGGAGAAAGCTGG   | CACCCACCCACTGGCCCCCTCAGA    | AGCGGGTGGCATA       | 240     |         |       |          |
| Qy                    | 306          | CTGCTTTGAGGTGGCAGCCAGCGAA   | GCCCAGATGAAAGACGTGCCCCAT    | TGAAGGAAGG          | 365     |         |       |          |
| Db                    | 241          | CTGCTTTGAGGTGGCAGCCAGCGAA   | GCCCAGATGAAAGACGTGCCCCAT    | TGAAGGAAGG          | 300     |         |       |          |
| Qy                    | 366          | AAACCCCTTTGGCCCATTTCTGGGAT  | CAGTTTCATGTGAGTTTCAACAAG    | TCGGAGCTTTT         | 425     |         |       |          |
| Db                    | 301          | AAACCCCTTTGGCCCATTTCTGGGAT  | CAGTTTCATGTGAGTTTCAACAAG    | TCGGAGCTTTT         | 360     |         |       |          |
| Qy                    | 426          | TACAGGCATTTCTTTCAGTGCTTCT   | TACAGAGAAACAATGGAGCCAGAGAT  | TTTCTCCAAA          | 485     |         |       |          |
| Db                    | 361          | TACAGGCATTTCTTTCAGTGCTTCT   | TACAGAGAAACAATGGAGCCAGAGAT  | TTTCTCCAAA          | 420     |         |       |          |

QY 486 GGAACATCCGCTGCTTGGCCCTGCCAGGAGCCCGAGCCAGTTCCTCCGCTCTAGAGACA 545  
Db 421 GGAACATCCGCTGCTTGGCCCTGCCAGGAGCCCGAGCCAGTTCCTCCGCTCTAGAGACA 480  
QY 546 CAGGCCACTACAGAGTACATGGTATGTGTACAGCAAAATGTGAAGACGGAGAGGCCCA 605  
Db 481 CAGGCCACTACAGAGTACATGGTATGTGTACAGCAAAATGTGAAGACGGAGAGGCCCA 540  
QY 606 GATTATGCCCACTGTGTGCGGCCCTTATGTGGGCATTCATCTGCGCATTTGGCTCTGACTG 665  
Db 541 GATTATGCCCACTGTGTGCGGCCCTTATGTGGGCATTCATCTGCGCATTTGGCTCTGACTG 600  
QY 666 GAAGAACGCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCTCTCC 725  
Db 601 GAAGAACGCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCTCTCC 660  
QY 726 GCAGTGTGTGGGCTACAGCGCAGACAGCGGCCCTCCACGATGACTATGTGCTGCC 785  
Db 661 GCAGTGTGTGGGCTACAGCGCAGACAGCGGCCCTCCACGATGACTATGTGCTGCC 720  
QY 786 TGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCTGCAGTC 845  
Db 721 TGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCTGCAGTC 780  
QY 846 GGTCTAGCTTGTACTGATTCGAGAGTATGTGCTGAGCTCCACAGCTCTCAAGG 905  
Db 781 GGTCTAGCTTGTACTGATTCGAGAGTATGTGCTGAGCTCCACAGCTCTCAAGG 840  
QY 906 GAAGTGAAGTGTGTAGCTTGAAGCTGAGTGTGGCCAGTGTGAGCTGTACATCTCGG 965  
Db 841 GAAGTGAAGTGTGTAGCTTGAAGCTGAGTGTGGCCAGTGTGAGCTGTACATCTCGG 900  
QY 966 CCAAGCGACACATTTATTTGGCAACTGTCTCTCTTCACTGTCTTTTGTGAAGCGGGA 1025  
Db 901 CCAAGCGACACATTTATTTGGCAACTGTCTCTCTTCACTGTCTTTTGTGAAGCGGGA 960  
QY 1026 GCGGACCTCCAGGGAGCGCTTCTTTCTTCGCGATGAGCAGGCCCTTAAGCTGCG 1085  
Db 961 GCGGACCTCCAGGGAGCGCTTCTTTCTTCGCGATGAGCAGGCCCTTAAGCTGCG 1020  
QY 1086 GGACGAGTTCTGATT 1100  
Db 1021 GGACGAGTTCTGATT 1035

RESULT 8

AAAD07629  
ID AAD07629 standard; cDNA; 915 BP.  
AC AAD07629;  
XX  
XX  
DT 10-AUG-2001 (first entry)  
XX Human secreted protein-encoding gene 18 cDNA clone HMWEY26, SEQ ID NO:69.  
XX Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; kidney disorder;  
KW cardiovascular disorder; angiotensin-related disorder; gene therapy;  
KW gastrointestinal disorder; pregnancy-related disorder; cell culture;  
KW endocrine disorder; infection; wound healing; vulnery; cell culture;  
KW chemotaxis; food additive; binding partner identification; chromosome 3;  
KW ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 1..498  
FT **1** \*tag= a  
FT **2** /product= "Human secreted protein precursor"

FT **3** /transl\_except= (pos:4..6, aa:Xaa)  
FT **4** /transl\_except= (pos:10..12, aa:Xaa)  
FT **5** /transl\_except= (pos:13..15, aa:Xaa)  
FT **6** /transl\_except= (pos:442..444, aa:Xaa)  
FT **7** /note= "Xaa equals any of the naturally occurring L-amino acids; CDS does not include start codon"  
FT **8** /partial  
FT **9** sig\_peptide 1..3  
FT **10** mat\_peptide 4..495  
FT **11** /\*tag= b  
FT **12** /\*tag= c  
FT **13** /product= "Mature human secreted protein"  
FT WO200132676-A1.  
FT 10-MAY-2001.  
FT 25-OCT-2000; 2000WO-US029365.  
FT 29-OCT-1999; 99US-0162237P.  
FT 21-JUL-2000; 2000US-0219666P.  
FT (HUMA-) HUMAN GENOME SCI INC.  
FT Ruben SM, Komatsoulis GA, Shi Y, Olsen HS, Soppet DR;  
FT WPI; 2001-328773/34.  
FT P-PSDB; AAE03110.  
FT Nucleic acids encoding 25 human secreted polypeptides, useful for preventing, diagnosing and/or treating e.g. Gaucher's disease, Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease, diabetes mellitus and multiple sclerosis.  
FT Claim 1; Page 445; 546pp; English.  
FT AAD07571-AA07645 represent cDNAs corresponding to 25 human secreted protein genes, and AAE03052-AAE03126 represent the proteins they encode. AAE03127-AAE03150 represent human secreted protein fragments. The genes and their corresponding secreted proteins are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 25 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiotensin-related disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present invention represents a human secreted protein-encoding cDNA of the invention  
FT SQ Sequence 915 BP; 209 A; 253 C; 247 G; 202 T; 0 U; 4 Other;  
FT Query Match 43.0%; Score 473.4; DB 4; Length 915;  
FT Best Local Similarity 98.4%; Pred. No. 8; 7e-126;  
FT Matches 477; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
FT QY 616 CACCTGTTCGGCCCTATGTGGCATTTCATCTGCGCATTCGCTCTGACTGGAGACGCC 675

```
Db 16 CACTTTGGCCGAACCTACCCGGGCAATTCATCTGGCAATGGCTCTGACTGGAAGAACGCC 75
Qy 676 TGTGCATCTGAAGGACGGAGCTGAGCTCGCACTTCATGGCTCTCCGCACTGCTG 735
Db 76 TGTGCATCTGAAGGACGGAGCTGAGCTCGCACTTCATGGCTCTCCGCACTGCTG 135
Qy 736 GGCTACAGCCGACGACAGCGGCCGCCCTCAGCATGACTATGTGCTCTGCTGACCTGAAG 795
Db 136 GGCTACAGCCGACGACAGCGGCCGCCCTCAGCATGACTATGTGCTCTGCTGACCTGAAG 195
Qy 796 GAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGATGCTGATGCTGCTGCTGCTG 855
Db 196 GAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGATGCTGATGCTGCTGCTGCTG 255
Qy 856 GCTACTGATTCGAGAGGTTATGTGCTGAGCTCCAAAGAGCTCTTCAAAGGGAAGGTGAAG 915
Db 256 GCTACTGATTCGAGAGGTTATGTGCTGAGCTCCAAAGAGCTCTTCAAAGGGAAGGTGAAG 315
Qy 916 GTGTGAGCCTGAAGCCTGAGGTGGCCCGAGGTGCACTGTATCTCTCGGCCAAGCCGAC 975
Db 316 GTGTGAGCCTGAAGCCTGAGGTGGCCCGAGGTGCACTGTATCTCTCGGCCAAGCCGAC 375
Qy 976 CACTTTATGGCACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1035
Db 376 CACTTTATGGCACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 435
Qy 1036 CAGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1095
Db 436 CAGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 495
Qy 1096 TGATT 1100
Db 496 TGATT 500
```

## RESULT 9

```
ADQ56824
ID ADQ56824 standard; DNA; 610 BP.
XX
AC ADQ56824;
XX
XX 21-OCT-2004 (first entry)
XX
DE Novel canine microarray-related DNA sequence SeqID8126.
XX
KW canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
OS Canis familiaris.
XX
XX WO2004063324-A2.
XX
XX 29-JUL-2004.
XX
XX 05-MAY-2003; 2003WO-US013853.
XX
XX 03-MAY-2002; 2002US-0377240P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX (PFIZ ) PFIZER PROD INC.
XX
XX Diggins JC, Porter M, Wei T;
XX
XX WPI; 2004-561890/54.
XX
XX New isolated nucleic acid molecule, useful for drug screening and
XX toxicity assays or for assessing the impact, including toxicity, of a
XX compound, pharmaceutical agent or environmental pollutant on a cell or
XX living organism.
```

Claim 1; SEQ ID NO 8126; 41pp; English..

This invention is related to a novel isolated canine nucleic acid sequences and the construction of canine microarrays containing a significant portion of the canine genome. The isolated canine nucleic acid sequences of the invention may be useful for drug screening and toxicity assays. The invention is therefore useful for assessing the impact, including toxicity, of a compound, pharmaceutical agent or environmental pollutant on a cell or living organism. The methods are useful for detecting genes that are up- or down-regulated in canines in a disease state. The sequences are useful as diagnostic agents or markers to detect a cellular response in a sample individually or as part of a gene expression profile. It is also useful as a target for agents that modulate gene expression or activity. The database is useful for producing electronic Northern blots that allow the user to determine the cell type or tissue in which a given gene is expressed and to allow determination of the abundance or expression level of a given gene in a particular tissue or cell. The methods are useful for determining the similarity of a toxic response to one or more individual compounds. The methods are useful for predicting at least one toxic response or the likelihood that a compound or test agent will induce various specific pathologies such as those of the liver (liver necrosis, fatty liver disease, protein adduct formation or hepatitis), those of the kidney, heart, brain or testes, or other pathologies associated with at least one of the toxins. The methods are also useful for predicting or elucidating the potential cellular pathways influenced, induced or modulated by the compound or test agent due to the similarity of the expression profile compared to the profile induced by a known toxin. The present sequence is that of a canine DNA sequence which was claimed for use during the production of a canine microarray of the invention.

Sequence 610 BP; 155 A; 171 C; 149 G; 129 T; 0 U; 6 Other;

Query Match 36.9%; Score 406.2; DB 13; Length 610;

Best Local Similarity 89.0%; Pred. No. 1.8e-106;

Matches 446; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

```
Qy 171 CACCAACTCCATGTGCTTACCAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCA 230
Db 1 CACCAATCTCCATGTGCTTACCAGAAAGTACTTCAAGCTNNAGCCCTCAGGNTACCA 60
Qy 231 TCGGCTCATCAGCTTGGAGGATTTTCATGGAAAGCTGGCACCCACCCACTCGCCCCCTGA 290
Db 61 TCGGTTTATCAGCTTAGAGACTTTCATGGAAGCTGGCACCCACTCACTGGCCCCCGA 120
Qy 291 GAAGCGGTGGCATACTGCTTTGAGTGGCAGCCAGGAGCCAGATAGAGACGTG 350
Db 121 GAGCGAGTGGCATACTGCTTTGAAGTGGCAGCCAGCCAGCCCTGATAGAGACATG 180
Qy 351 CCCCATGAAGGAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAA 410
Db 181 CCCCATGAAGGAAGAAATCCCTTTGGCCCATTTTGGATCAGTTTCATGTGAGTTTCAA 240
Qy 411 CAAGTCGGAGCTTTTTCAGGCAATTCCTTCAGTGTCTTCTACAGAGAACATGAGCCA 470
Db 241 CAAGTCCAAAGCTTTTTCAGGCAATTCCTTCAGTGTCTTCTACAAAGACCAATGGATTCA 300
Qy 471 GAGATTTTCTCAAAGGAACATCCGGTGTCTTGGCTTGGCCAGGAGCCAGCCAGTTCCC 530
Db 301 GAGATTTTCTCAAAGGAACATCCAGTGTCTTGGCCCTCCAGGAGCCAGCCAGTTCCC 359
Qy 531 CGTCTAGAGAACACAGGCGCACTACAGAGTACATGTTGTTGTTGTTGTTGTTGTTGTTGTT 590
Db 360 CGTCTAGAGAACACAGGCGCACTACAGAGTACATGTTGTTGTTGTTGTTGTTGTTGTTG 419
Qy 591 GACGGAGAGGCCAGATTCATGCCACCTTGTCCGGCCCTTATGTGGGCAATTCATCTCGG 650
Db 420 GACGGAGAGGCCAGATTCATGCCACCTTGTCCGGCCCTTATGTGGGCAATTCATCTCGG 479
Qy 651 CATTGGCTCTGACTGGAGAA 671
Db 480 CATTGGCTCTGACTGGATGGA 500
```

RESULT 10  
AAK79915  
ID AAK79915 standard; DNA; 10331 BP.  
XX  
AC AAK79915;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34727.  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205115P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 25-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0225219P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 08-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 28-SEP-2000; 2000US-0236327P.  
PR 28-SEP-2000; 2000US-0236367P.  
PR 28-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.



PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234237P.  
PR 25-SEP-2000; 2000US-0234397P.  
PR 25-SEP-2000; 2000US-0234398P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 25-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249254P.  
PR 17-NOV-2000; 2000US-0249255P.  
PR 17-NOV-2000; 2000US-0249256P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders, including a  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX SQ Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;  
Query Match 22.8%; Score 251.2; DB 4; Length 10331;  
Best Local Similarity 97.0%; Pred. No. 3.4e-61;  
Matches 256; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 656 GCTCTGACTGGAAGACGCTGTGCCATGCTGAAGGACGGGACTGACAGCTCGCACTTCA 715  
Db |||||  
QY 716 TGGCTCTCCGAGTGTGGCTACAGCCGACGACAGGGCCCCCTCAGCATGACTA 775  
Db |||||  
QY 776 TGTGCTGCTGACTGAGAGATCCAGAGGCTGTGAAGCTCTGGTGAAGTCCGCTGG 835  
Db |||||  
QY 836 ATGCCAGTCCGCTCTACGTTGCTACTGATCCGAGATTATGCTGAGCTCCACAGC 895  
Db |||||  
QY 896 TCTTCAAAGGGAAGGTGAAGTGG 919  
Db |||||  
ID ACH32225 standard; cDNA; 479 BP.  
XX ACH32225;  
XX 13-OCT-2003 (first entry)  
XX Human endothelial cell cDNA #358.  
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
XX genome mapping; biodiversity; genetic disorder.  
XX Homo sapiens.  
XX US2003073623-A1.  
XX 17-APR-2003.  
XX 30-JUL-2001; 2001US-00918995.  
XX



```

PR 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 19437; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 479 BP; 99 A; 140 C; 143 G; 97 T; 0 U; 0 Other;
XX
XX Query Match 21.2%; Score 233; DB 9; Length 479;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-56;
XX Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 868 GAGAGTTATGTCCTGAGCTCCAGCTCCACAGCTTTCAAAGGGAAGGTGAGTGGTGGCTG 927
DB |||||||
DB 31 GAGAGTTATGTCCTGAGCTCCAGCTCCACAGCTTTCAAAGGGAAGGTGAGTGGTGGCTG 90
QY 928 AAGCCTGAGGTGGCCAGGTGCACTGTGACATCTCTCGGCCAAGCGCACCTTTATTGGC 987
DB |||||||
DB 91 AAGCCTGAGGTGGCCAGGTGCACTGTGACATCTCTCGGCCAAGCGCACCTTTATTGGC 150
QY 988 AACTGTGTCTCTCTTCACTGCTTTGTGAGCGGAGCGGACCTCCAGGGAGGCCG 1047
DB |||||||
DB 151 AACTGTGTCTCTCTTCACTGCTTTGTGAGCGGAGCGGACCTCCAGGGAGGCCG 210
QY 1048 TCTTCTTTCTTCGGCATGCAAGCGCCCTTAAGCTGCGGACGAGTTCCTGATT 1100
DB |||||||
DB 211 TCTTCTTTCTTCGGCATGCAAGCGCCCTTAAGCTGCGGACGAGTTCCTGATT 263

RESULT 13
ACH35532
ID ACH35532 standard; cDNA; 477 BP.
XX
XX ACH35532;
AC
XX 13-OCT-2003 (first entry)
XX
XX Human endothelial cell cDNA #3665.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX

```

```

OS Homo sapiens.
XX
XX US2003073623-A1.
XX
XX 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 22744; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 477 BP; 95 A; 141 C; 142 G; 97 T; 0 U; 2 Other;
XX
XX Query Match 20.2%; Score 221.8; DB 9; Length 477;
XX Best Local Similarity 96.3%; Pred. No. 2.5e-53;
XX Matches 237; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
XX
QY 855 TGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCACAGCTCTTCAAAGGAAGGTGAA 914
DB |||||||
DB 11 TGATGCTGAGCNCCTAGAGATATGTGCTGAGCTCC-ACAGTTCTTCAAAGGAAGGTGAA 69
QY 915 GGTGCTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATCTCTCGGCCAAGCCGA 974
DB |||||||
DB 70 GGTGCTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATCTCTCGGCCAAGCCGA 129
QY 975 CCACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAGCGGAGCGGACCT 1034
DB |||||||
DB 130 CCACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAGCGGAGCGGACCT 189
QY 1035 CCAGGGAGGCCGCTCTTCTTTCTTCGGCATGGAAGGCCCTTAACTGTCGGGACAGTT 1094
DB |||||||
DB 190 CCAGGGAGGCCGCTCTTCTTTCTTCGGCATGGAAGGCCCTTAACTGTCGGGACAGTT 249
QY 1095 CTGATT 1100
DB |||||||
DB 250 CTGATT 255

RESULT 14

```

ABL13947  
ID ABL13947 standard; cDNA; 1209 BP.  
XX  
AC ABL13947;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
XX  
DR P-PSDB; ABB69844.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;  
Query Match 15.1%; Score 166.4; DB 4; Length 1209;  
Best Local Similarity 51.5%; Pred. No. 3.7e-37;  
Matches 563; Conservative 0; Mismatches 486; Indels 45; Gaps 6;  
17 GGGACCCGCGGTACTGCTCTACTGCTCCCTGATGGGCGCTTTGGGAACCGAGCGG 76  
80 GCGATCCCAATGGCTACTCACTACTGCTCGGTATGGAGCGCTTTGGCAACCGAGCGG 139  
77 ATCACTTCTGGGCTCTGGGATTTGGAAGCTCTAAACCTGCTTGGCTGCTCCCTC 136  
140 ACCACTTCTGGGATCATTTGGCTTCCGCAAGCGCTTATCGACCTGATCTGCTGCGC 199  
137 CTGTGATTTAGTACGAGCATCAAGCTCTTTCCCAACCTCCATGCTCTACACAGA 196  
200 CGTGGGTGGAGT-----ATCGTAGGGGTGAACCTGCTACCTGCTACCTGCTACACA 253  
197 AGTACTTCAAGCTGAGGCCCTCCAGGCTTACCATCGGCTCATCGCTTGGAGGATTTCA 256  
254 CATATTTTGAAGTGAGGCCCTGAAGGAATACCATCGGCTCATCAGGATTTCA 313  
257 TGGAGAAGCTGGCAACCCACCTCGGCCCTCGAGAAGCGGGTGGCATACTGCTTTGAGG 316  
314 TGTGGACCTGGCCGACGACATTTGGCCGAATTCGAGCGAGTGTCTATTTTGTACAGG 373

QY 317 -----TGGCAGCCCGAGCGAAGCCAGATAGAAGACGTGCCCCATGAAGG 361  
DB 374 AACGATATAGCTTTCAGCAGGAGAAAGACGATCCAGAACAGCCAAATGCCAGCGCAAGG 433  
QY 362 AAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTAGTTTCAACAAGTCGGAGC 421  
DB 434 ATGGCAATCCCTTTTGGTCCCTTTTGGGACACTTTTTCATAGACTTTTGTGCGGTGAGAGT 493  
QY 422 TTTTTCACAGGCAATTTCTTCAGTCTTCTCTACAG-----AGAACATGAGGCCAGA 472  
DB 494 TCTATGCGCCACTTTCATTTTGAATGTGATCATATAGCAACAGAGGCTGCCAAGTGGCAGACCA 553  
QY 473 GATTTTCTCCAAAGGAACATCCGGTGTGCTGCGCTGCGAGGAGCCCGAGCCAGTTTCCCGC 532  
DB 554 AATATCTGCGAGATCATATCCGTTACTCGGTTTCAACGAGGCTCCGGCTAGTTTTCCTG 613  
QY 533 TCCTAGAAGAACACAGCGCACTACAGAGTACATGTTATGTCAGTGAAGTCAACGGTATAGGAG 592  
DB 614 TTACGTAGAACTGCAAGCTGACGCTACTTTCAGTGAAGTCAACGGTATAGGAG 673  
QY 593 CGGAGAGGCCAGATTTCATGCCACCTTGTCCCGC-----CCTATGTGGGATTCATCTGC 649  
DB 674 CATCTAAGGATTTTCATCCGAGACGTTGCTTGGGGTGCCTTTTGGGCAITTCATCTGC 733  
QY 650 GCATTGGCTCTGACTGGAAGAACGCTGTGCCATGCTGAAGACGG-----GACTG 700  
DB 734 GCAACGGTATCGATTGGGTGAGAGCCTGTGAGCAGCTCAAGGATAGCCAGCATCTGTGTTG 793  
QY 701 CAGGCTCGCACTTTCATGGCCTCTCCGAGTGTGGGCTACAGCCGAGCAGCAGCGGCC 760  
DB 794 CTTCCGCGCAGTGTCTGGGCTATAAATAAGAGTGTGTCACCTTACCCGAGCTCTGCA 853  
QY 761 CCTCAGCATGACTATGTGCTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT 820  
DB 854 TGCCTTCCAGGAGCGGATCATCCGCACTAAGAGAAACCTTAAGAACGTGCGCCAAA 913  
QY 821 GGGTGAAGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 880  
DB 914 CTCAGCCGACAAAGAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 973  
QY 881 CTGAGCTCCACAGCTCTTCAAGGAGAGTGAAGGTGCTGAGCTGAGCTGAGCTGAGCTGAG 940  
DB 974 GTGACTAAACACGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1033  
QY 941 ---CCAGGTGACCTGTATCATCTCGGCAAGCCGACACTTTTATTTGGCACTGTGCT 997  
DB 1034 ATCTTTACTGAGCTTGGCCATTTCTCGACAGTCAAGCACTTTATCGCACTGTATAT 1093  
QY 998 CCTCTTCACTGCTCTTTTGAAGCGGAGCGGACCTCCAGGGAGCGCTCTTCTTCT 1057  
DB 1094 CTTCTTACTCGGCATTCGAAAAAAGGGACGAGATGTGCACGCTTTTCCATCGTACTTCT 1153  
QY 1058 TCGGCATGAGCAGG 1071  
DB 1154 GGGGATTTCCCAAG 1167  
RESULT 15  
ABL13946/C  
ID ABL13946 standard; cDNA; 3264 BP.  
XX  
AC ABL13946;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.

XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX P-PSDB; ABB69843.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
XX ABBS72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;  
Query Match 13.7%; Score 150.4; DB 4; Length 3264;  
Best Local Similarity 50.9%; Fred. No. 2.4e-32;  
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;  
Qy 49 TGCATGGGGCGCTTTGGAAACAGGCCGATCACATCTCTGGGCTCTCTGGCATTTGGCAAAG 108  
Db TTTCAAGGACGCTTTGGCAACAGGCCGACCATCTCTGGGATCATTTGGCCCTTGGCCAG 2039  
Qy 109 CTGCTAAACCGTACTTGGCTGTCCCTCTTGGATTGAGTACCAGCATCAAGCCCTCT 168  
Db CGCTTTAATCGACCCCTGATCTCGCCGCTGGGTGGAGT-----ATCGTAGGGTGAA 1985  
Qy 169 TTCACCAACCTCCATGTCTCTACAGAACTTCAAGTCTTCAAGTCTGGAGCCCTCCAGCTTAC 228  
Db CTGGATCCCGTCAAGTACCCTTCAACACATATTTTGAAGTGGAGCCCTGAAGGAATAC 1925  
Qy 229 CATCGGTCATCAGCTTGGAGGATTTATGAGAACTGGGACCCACCCACTGGCCCTCT 288  
Db CATCGGTCATCACCATTGGAGATTTATGAGAACTGGGACCCCTGGGAGGACATTTGGCCAGAA 1865  
Qy 289 GAGAAAGCGGTGGCATACTGCTTTTAGG-----TGGCAGCCCAAGCGAAGC 333  
Db TCGGAGCGAGTGTATTTTGTCTAAGGAACGATATAGCTTTCAGCAGGAGAAAGCAT 1805  
Qy 334 CCAGATAAGAGACGTGCCCCATGAAGAAAGAACCCCTTTGGCCCATCTGGGATCAG 393  
Db CAGACAAAGCCCAATTTGGCAGCAAGGATGGCAATCTTTTGGTCCCTTTGGGCACT 1745  
Qy 394 TTTTCATGTGAGTTTCAACAACTGGAGCTTTTACAGGCAATTTCTTTCAGTCTTCTAC 453  
Db TTTTCATAGACTTTTGGCGGTGAGATTTCTATGGGCACTTTTATTTGATGTGCATCAT 1685  
Qy 454 AG-----AGAACATATGGAGCCAGAGATTTTCTCAAGAAACATCCGGTCTTGGCC 504  
Db AGCAACGAGGCTGCCAAGTGGCAGACCAAAATATCTTCAGAAATCATATCCCGTACTCGCG 1625  
Qy 505 CTGCGAGGAGCCCGAGCCCGCTTCCCGCTCTAGAAAGAACACAGGCCCATACAGAACTAC 564  
Db TTCACCGAGGCTCGGGCTAGTTTCTCTGCTAGAGAACTGCAAGCTGCAGCGCTAC 1565

Qy 565 ATGGTATGTCAGACGAAATGGTGAAGACGGAGAGGCCAGATTTCATGCCCCACCTTGTTC 624  
Db TTTTGGAGTGGAGTCAACGGTATAGGAAGCATCTAAGGATTTTCATCCGAGAGGATGGCT 1505  
Qy 625 CGG---CCCTATGTGGGCATTCATCTGCGCATTTGGCTCTGACTGGAAGAACGCTTGTGCC 681  
Db TTTTGGGTCCTTTTGGGTCATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTTGTGAG 1445  
Qy 682 ATGCTGAGGAGCGG-----GACTGCGAGGCTCGCACTTTCATGGCCTCTCCGCAAGTGT 732  
Db TTTTGGGTCATGAGGATAGCCAGCATCTGTGCTTCGCGCAGTGTGTTGGGCTATAAAAAATGAA 1385  
Qy 733 GTGGGCTACAGCCGAGCAGCAGCGGCCCTCCCTCAGCATGACTATATGTGCTGTGCTGACCTG 792  
Db TTTTGGTGTGCTTACCCGAGCTCTGCAATGCTCCCTCAAGGAGGCGCATCATCCGCGAGCTA 1325  
Qy 793 AAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCGCCAGTCGCTGTCTAC 852  
Db TTTTGAAGAACCATTAAGAACGTCGCGCAAACTCAGCGCGGACAAACGAAATCAAAATCAGTTTC 1265  
Qy 853 GTTCTACTGATTCGAGAGTATATGTGCTGAGCTTCCACAGCTTTCACAGGGAAGGTG 912  
Db TTTTGTGGCTCAGACTCCAATTCACATGATTTGGTGAACCTAAACACGCGCCCTTAGTGCATGGGC 1205  
Qy 913 AAGGTGGTGGCTTGAAGCCCTGAGGTGG---CCGAGGTGCGACCTGTACATCTCTCGGCCAA 969  
Db TTTTATCATGTGCAACAAGCTGCGGAGGATGATCTTACCTGGACTTGGCCATTTCTCGGACAG 1145  
Qy 970 GCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAAGCGGAGCGG 1029  
Db TTTTTCGAACCACTTTATTCGCGCACTGTATATCTCTTACTCGGCATTCGAAAAAAGGGAACGA 1085  
Qy 1030 GACCTCCAGGAGGCGCTCTTCTTTCTTCTTGGCATGACAGG 1071  
Db TTTTATGTCACGCTTTTTCATCTGCTTCTTGGGATTCGCCAAG 1043

Search completed: November 9, 2005, 04:43:20  
Job time : 644.158 secs

THIS PAGE BLANK (USPTO)

| Result No. | Query  |       |        | DB | ID        | Description |
|------------|--------|-------|--------|----|-----------|-------------|
|            | Score  | Match | Length |    |           |             |
| 1          | 1096.8 | 99.7  | 1555   | 3  | BR617950  | full-length |
| 2          | 892.8  | 81.2  | 1141   | 5  | BR353274  | BR353274    |
| 3          | 885.4  | 80.5  | 3611   | 3  | AK081059  | Mus muscu   |
| 4          | 884.2  | 80.4  | 2616   | 3  | AK044629  | Mus muscu   |
| 5          | 743.2  | 67.6  | 1091   | 5  | BM909055  | AGENCOURT   |
| 6          | 718.4  | 65.3  | 841    | 5  | BR3435989 | BR3435989   |
| 7          | 717.4  | 65.2  | 719    | 1  | AL135434  | DKF297620   |
| 8          | 712.2  | 64.7  | 889    | 4  | BI524031  | BI524031    |
| 9          | 706.6  | 64.2  | 898    | 5  | BR349372  | BR349372    |
| 10         | 673.2  | 61.2  | 942    | 7  | CO774833  | ILLUMIGEN   |
| 11         | 662.2  | 60.2  | 808    | 4  | BG747760  | BG747760    |
| 12         | 660.2  | 60.0  | 801    | 5  | BP144264  | BP144264    |
| 13         | 639    | 58.1  | 988    | 5  | BQ964179  | AGENCOURT   |
| 14         | 636.4  | 57.9  | 874    | 4  | BI757540  | BI757540    |
| 15         | 630.8  | 57.3  | 947    | 4  | BI410783  | BI410783    |
| 16         | 625.6  | 56.9  | 809    | 4  | BG473023  | BG473023    |
| 17         | 624    | 56.7  | 915    | 5  | BQ936634  | BQ936634    |
| 18         | 609.8  | 55.4  | 705    | 7  | CN348925  | CN348925    |
| 19         | 604.6  | 55.0  | 919    | 5  | BU155142  | AGENCOURT   |
| 20         | 599.2  | 54.5  | 753    | 4  | BG424731  | BG424731    |
| 21         | 578.2  | 52.6  | 689    | 4  | BI559635  | BI559635    |
| 22         | 558.4  | 50.8  | 561    | 4  | BM798261  | BM798261    |
| 23         | 542.4  | 49.3  | 760    | 5  | BU055535  | UI-N-FOO    |
| 24         | 541.8  | 49.3  | 545    | 7  | CN348928  | CN348928    |

Db 109 ATGCGCTGCGGGCTCTTGGGAGCCGGCCGGTTACTGCTCTACTGCCCCCTGCANTGGGGGCG 168  
Qy 61 TTTGGGAACAGGCGCGATCACTTTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 120  
Db 169 TTTGGGAACAGGCGCGATCACTTTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 228  
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTTAGTTACAGATCACAAGCTCTCTTTACCAACCTC 180  
Db 229 ACCTTGGCTGTCCCTCTCTGGATTTAGTTACAGATCACAAGCTCTCTTTACCAACCTC 288  
Qy 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 289 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 348  
Qy 241 ACCTTGGAGGATTTTCATGGAGAGCTGGGACCCACCCTCTGGGCCCCCTGAGAACGGGGT 300  
Db 349 ACCTTGGAGGATTTTCATGGAGAGCTGGGACCCACCCTCTGGGCCCCCTGAGAACGGGGT 408  
Qy 301 GCATACCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGCCCCATGAAG 360  
Db 409 GCATACCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGCCCCATGAAG 468  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 469 GAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 528  
Qy 421 CTTTTCAGGCGATTTCTTCTAGTCTTCTTCTACAGAAACAATGGAGCCAGAGATTTTCT 480  
Db 529 CTTTTCAGGCGATTTCTTCTAGTCTTCTTCTACAGAAACAATGGAGCCAGAGATTTTCT 588  
Qy 481 CCAAGGAACATCCGGTGTCTTGGCTTGGCCAGGAGCCCGCCAGTTCCCGTCTCTAGAA 540  
Db 589 CCAAGGAACATCCGGTGTCTTGGCTTGGCCAGGAGCCCGCCAGTTCCCGTCTCTAGAG 648  
Qy 541 GAAACAGGCGCACTACAGAGTACATGTGTATGGTCAGACGAATGTGAAGACGGGAGAG 600  
Db 649 GAAACAGGCGCACTACAGAGTACATGTGTATGGTCAGACGAATGTGAAGACGGGAGAG 708  
Qy 601 GCCAGATTCATGCGCCACTTGTGCGGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 660  
Db 709 GCCAGATTCATGCGCCACTTGTGCGGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 768  
Qy 661 GACTGGAAGAACGCTGTGTCATGCTGAAGACGGGATCTGAGGCTCGCACTTCATGCGC 720  
Db 769 GACTGGAAGAACGCTGTGTCATGCTGAAGACGGGATCTGAGGCTCGCACTTCATGCGC 828  
Qy 721 TCTCCGCACTGTGGGCTACAGCCGACAGCAGCGGCCCTCCACGATGACTATGTGC 780  
Db 829 TCTCCGCACTGTGGGCTACAGCCGACAGCAGCGGCCCTCCACGATGACTATGTGC 888  
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGCC 840  
Db 889 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGCC 948  
Qy 841 CAGTGGCTTACGTTGCTACTGATTCAGAGGTTATGTGGCTGAGCTCCACAGCTCTTC 900  
Db 949 CAGTGGCTTACGTTGCTACTGATTCAGAGGTTATGTGGCTGAGCTCCACAGCTCTTC 1008  
Qy 901 AAAGGAAGGTGAAGGTGAGCTGAAGCTGAGGTGGCGCCAGGTGCACTGTGATCATC 960  
Db 1009 AAAGGAAGGTGAAGGTGAGCTGAAGCTGAGGTGGCGCCAGGTGCACTGTGATCATC 1068  
Qy 961 CTCGCCCAAGCCGACCACTTTATTTGSCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
Db 1069 CTCGCCCAAGCCGACCACTTTATTTGSCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1128  
Qy 1021 CGGGAGCGGAGCTTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080  
Db 1129 CGGGAGCGGAGCTTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1188  
Qy 1081 CTGCGGAGCGAGTTCTGATT 1100  
Db 1189 CTGCGGAGCGAGTTCTGATT 1208

RESULT 2  
BX353274  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BX353274 1141 bp mRNA linear EST 08-APR-2004  
BX353274 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
cDNA clone CS0DC003YJ11 5-PRIME, mRNA sequence.

BX353274  
BX353274.2 GI:46290116  
EST.

Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 1141)

Li.W.B., Gruber.C., Jessee.J. and Polayes.D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 5, 2003 this sequence version replaced gi:30371746.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

9980.f

For more information about this cluster, see

<http://www.genoscope.cns.fr/cdna?s=CS0DC003CS060P1&c=9980.f>.

Location/Qualifiers

1. 1141

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="CS0DC003YJ11"

/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 81.2%; Score 892.8; DB 5; Length 1141;  
Best Local Similarity 92.8%; Pred. No. 8.9e-235;  
Matches 952; Conservative 32; Mismatches 35; Indels 7; Gaps 5;

Qy 1 ATGCCCCGGGGCTCTGGGACCCGGCGGGTTACCTGCTCTACTGCCCCCTGCATGGGGGCG 60

Db 86 ATGCCCCGGGGCTCTGGGACCCGGCGGGTTACCTGCTCTACTGCMCTGCATGGGGGCG 145

Qy 61 TTTGGGACACGAGCCGATCACTTTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 120

Db 146 TTTGGGAACAGGCGCGATMACTTTTGGGCTCTTGGCATTTGCAAGCTGCGAAMWGT 205

Qy 121 ACCTTGGCTGTCCCTCTTGGATTGATTACAGCATCACAAGCTCTCTTTCACCAACCTC 180

Db 206 ACCTTGGCTGTACCTCTCTTGGATTGATTACAGCATCAMAAGCTCTTTCACCAACCTC 265

Qy 181 CATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240

Db 266 CATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 325

Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCACTGGGCCCCCTGAGAAGCGGGT 300

Db 326 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCACTGGGCCCCCTGAGAAGCGGGT 385

Qy 301 GCATACCTCTTGGAGTGGCAGCCAGGAGCCAGATGAAGAGAGCTGCCCATGAAG 360

Db 386 GCATACCTCTTGGAGTGGCAGCCAGGAGCCAGATGAAGAGAGCTGCCCATGAAG 444







## JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: <http://genome.gsc.riken.jp/>; Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

cdNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Retina RNA was provided by Dr. Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged. Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>

FEATURES  
source

```

1. .2616
   Location/Qualifiers
     /organism="Mus musculus"
     /mol_type="mRNA"
     /strain="C57BL/6J"
     /db_xref="FANTOM_DB:A930028F21"
     /db_xref="taxon:10090"
     /clone="A930028F21"
     /tissue_type="retina"
     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     /dev_stage="adult"
34. .1215
   /note="unnamed protein product; protein
   O-fucosyltransferase 1 (MGD|MGI:2153207, GB|NM_080463,
   evidence: BLASTN, 99%, match=1290)
   putative"
   /codon_start=1
   /protein_id="BAC32009.1"
   /db_xref="GI:26336653"
   /translation="MGAAPAPHLRLRSLFLLLLPLRGRSAGSWDLAGLYLLCP
   CMGRGNADHFLGSLAKILNRTLAVPMLEVOHKKPPTNLHVSQYKLEPLQ
   AYHRVSLDEPMENLAPSHWPEKRVAYCFEVAQRSPDKTCEPKNGNPRGFWDQF
   HVSFNSELTGIFSASVKGQWTFORFAKEHPVLALPGAQFPVLEHRELOKYMV
   WSDMVRTGEIISHLVYGVYTHLRIGSDWKNACMLKDTAGSHFPMASPOCVGYS
   RSTAPLMTMCLPDKIEIORAVTLWRLNARSVYIATDSYSVSEIQQLFKDKVRV
   VSLKEVAQIDLYLQADHFIGNCVSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
   F"

```

## CDS

```

/note="unnamed protein product; protein
O-fucosyltransferase 1 (MGD|MGI:2153207, GB|NM_080463,
evidence: BLASTN, 99%, match=1290)
putative"
/codon_start=1
/protein_id="BAC32009.1"
/db_xref="GI:26336653"
/translation="MGAAPAPHLRLRSLFLLLLPLRGRSAGSWDLAGLYLLCP
CMGRGNADHFLGSLAKILNRTLAVPMLEVOHKKPPTNLHVSQYKLEPLQ
AYHRVSLDEPMENLAPSHWPEKRVAYCFEVAQRSPDKTCEPKNGNPRGFWDQF
HVSFNSELTGIFSASVKGQWTFORFAKEHPVLALPGAQFPVLEHRELOKYMV
WSDMVRTGEIISHLVYGVYTHLRIGSDWKNACMLKDTAGSHFPMASPOCVGYS
RSTAPLMTMCLPDKIEIORAVTLWRLNARSVYIATDSYSVSEIQQLFKDKVRV
VSLKEVAQIDLYLQADHFIGNCVSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
F"

```

## ORIGIN

```

Query Match      80.4%; Score 884.2; DB 3; Length 2616;
Best Local Similarity 87.9%; Pred. No. 2.6e-232;
Matches 964; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 3  GCCCGGGGCTCTGGGACCGCGGTTACTGCTCTACTGCCCTCGATGGGGCGTT 62
Db 120  GTCAGCGGGCTCTGGGACCTGGCGGTTACCTGCTCTACTGCTCTGATGGGGCGTT 179

Qy 63  TGGGAACAGCGCGATCTCTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGTAC 122
Db 180  TGGGAACAGCGCTGATCACTCTTGGGCTCTGCGCATTTGGAAGCTGCTGAACCGCAC 239

Qy 123  CTTGGCTGTCCCTCTGGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 182
Db 240  CTTGGCTGTACCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 299

Qy 183  TGTGTCTTACAGAAAGTACTTCAAGCTGGAGCGCCCTCCAGGCTTACCATCGGGTCAATCAG 242
Db 300  TGTGTCTTACAGAAAGTACTTCAAGCTGGAGCGCTCTCCAAAGCTTACCATCGGGTGTGAG 359

Qy 243  CTTGAGGATTTTCATGGAGAGCTGGCACCACCTGCGCCCTGGAGAGCGGGTGGC 302
Db 360  CTTGAGGATTTTCATGGAGAGCTGGCACCCTCCACCTGGCGCCCTGGAGAGCGGGTGGC 419

Qy 303  ATACTGCTTTGAGTGGGAGCGCCAGGAGCCAGATAAGAGAGCTGCCCATGAAGA 362

```

```

Db 420  ATACTGCTTTGAGTGGGAGCGCCAGCAAGTCTCTGATAAGAGACATGTCCTCATGAAGA 479
Qy 363  AGGAAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTCAGTTTCAACAAGTCGGAGCT 422
Db 480  AGGAAATCCTTTTGGCCCATTTCTGGACACAGTTTCTGTCAGTTTCAATAAGTCAGAACT 539
Qy 423  TTTTACAGGCATTTCTTTCAGTGTCTCTACAGAGAACAAATGGAGCCAGAGATTTTCTCC 482
Db 540  GTTCACAGCATTTCTTTCAGCGCTCTCTACAAAGAACAAATGGAGCCAGAGATTTCTTCC 599
Qy 483  AAAGAACATCGGTGCTTGGCCCTGCGCAGGAGCCCGCCAGCCAGTTTCCCGCTCTCTAGA 542
Db 600  AAAAGAGATCTCTGTCGCACTGCTGCGGCGCCCGCCAGCAGTTCCTTGTCTCTGAGAGA 659
Qy 543  ACACAGGCCACTACAGAGATGATGATGTCAGACCAAAATGGTGAAGACGGAGAGCG 602
Db 660  ACACAGGAGCTCCAGAAATGATGATGTCAGATGATGATGATGATGATGATGATGATGATG 719
Qy 603  CCAGATTCATGCCCACTTGTCCGCGCCCTATGTCGCGCATTTCTGCGCATTTGGCTCTGA 662
Db 720  CCTGATCAGTTCCCACTTGTCCGCGCCCTATGTCGCGCATTTCTGCGCATTTGGCTCTGA 779
Qy 663  CTGGAAGAACCTGTCGTCATGTCGAGAGCGGAGCTGAGAGCTGCGACTTTCATGCGCTC 722
Db 780  CTGGAAGAAATGCTGTCGTCATGTCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 839
Qy 723  TCCGAGTGTCTGGGCTACAGCCGACGACGAGCGGCGCCCTCTGAGTGAAGTGAAGTGAAG 782
Db 840  CCCTCAGTGTGGGCTATAGCCGACGACGAGCGGCGCCCTCTGAGTGAAGTGAAGTGAAG 899
Qy 783  CCCTGACCTGAAGAGATCCAGAGCGGCTGTGAAGCTCTGGGTGAGGTGAGTGGATGCCCA 842
Db 900  CCCTGACCTGAAGAAATCCAGCGGCTGTGAAGCTCTGGGTGAGAGTGAAGTGAAGTGAAG 959
Qy 843  GTCGCTCTACTGCTCTGATTCGAGAGTATGTCGCTGAGTCTCAACAGCTCTTCAA 902
Db 960  ATCGGCTTACATCGCCACAGACTCTGAGAGTACTGTCGAGAGATCCAGAGCTCTTCAA 1019
Qy 903  AGGGAAGTGAAGTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 962
Db 1020  AGAAGAAGTGAAGTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1079
Qy 963  CGGCAAGCCCAACCACTTATTTGCAACTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1022
Db 1080  CGGCAAGCTGACCACTTATTTGCAACTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1139
Qy 1023  GGAGCGGACCTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1082
Db 1140  GGAGCGGACCTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1199
Qy 1083  CGGGAAGCTGTCGAT 1099
Db 1200  TCGGATGAATTTGAT 1216

RESULT 5
BM909055 1091 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT 6611915 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431669
5', mRNA sequence.
ACCESSION BM909055
VERSION BM909055.1 GI:19359434
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1091)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

```

Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L10M1904 row: e column: 14  
High quality sequence stop: 629.  
Location/Qualifiers  
1..1091  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5431669"  
/tissue\_type="astrocytoma grade IV, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 98"  
/note="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGACAGG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

FEATURES  
source

## ORIGIN

Query Match 67.6%; Score 743.6; DB 5; Length 1091;  
Best Local Similarity 93.0%; Pred. No. 1.3e-193;  
Matches 824; Conservative 0; Mismatches 54; Indels 8; Gaps 4;

Qy 178 CTCATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGAGCTTACCATCGGTC 237  
Db 132 CTCATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGAGCTTACCATCGGTC 191

Qy 238 ATCAGCTTGGAGGATTCATGAGAGCTGGACCCACCTGCGCCCTCAGAGCGG 297  
Db 192 ATCAGCTTGGAGGATTCATGAGAGCTGGACCCACCTGCGCCCTCAGAGCGG 251

Qy 298 GTGGCATACTCTTTGAGTGGAGCCAGGAGCCAGATGAAGACGTGCCCATG 357  
Db 252 GTGGCATACTCTTTGAGTGGAGCCAGGAGCCAGATGAAGACGTGCCCATG 311

Qy 358 AAGGAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 417  
Db 312 AAGGAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 371

Qy 418 GAGCTTTTACAGGATTTCTTCACTGCTTCCTACAGAGCAATGAGCCAGAGATTT 477  
Db 372 GAGCTTTTACAGGATTTCTTCACTGCTTCCTACAGAGCAATGAGCCAGAGATTT 431

Qy 478 TCTCAAGAGAACATCCGGTGTGCTTGGCTTGCAGAGAGCCAGCCAGTTCCCGTCTTA 537  
Db 432 TCTCAAGAGAACATCCGGTGTGCTTGGCTTGCAGAGAGCCAGCCAGTTCCCGTCTTA 491

Qy 538 GAAGAACACAGGCCCACTACAGAGTACATGTGTGTCAGACGAAATGGTGAAGACGGGA 597  
Db 492 GAAGAACACAGGCCCACTACAGAGTACATGTGTGTCAGACGAAATGGTGAAGACGGGA 551

Qy 598 GAGGCCAGATTTATGTCACCTTGTTCGGCCCTTATGTGGGCAATTCATCTGGCATTGGC 657  
Db 552 GAGGCCAGATTTATGTCACCTTGTTCGGCCCTTATGTGGGCAATTCATCTGGCATTGGC 611

Qy 658 TCTGACTGGAAGACCGCTGTGCTATGTGAGAGCGGACTCAGGCTCCGATTCATG 717  
Db 612 TCTGACTGGAAGACCGCTGTGCTATGTGAGAGCGGACTCAGGCTCCGATTCATG 671

Qy 718 GCCTCTCCGCATGTGTGGGCTTACAGCGCAGCACAGCGGCCCTCCCTACGATGACTATG 777  
Db 672 GCCTCTCCGCATGTGTGGGCTTACAGCGCAGCACAGCGGCCCTCCCTACGATGACTATG 731

Qy 778 TGCCTCGCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTGCCTGGAT 837

Db 732 TGCCTCGCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTGCCTGGAT 791

Qy 838 GCCAGTCCGGTCTACGTGTGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTC 897

Db 792 GCCAGTCCGGTCTACGTGTGCTACTGATTCCGAGAGTTATGTGCTGAAAGCTCCACAGCTC 851

Qy 898 TTCAAA--GGGAAGGTGAAGTGGTG--AGCTGAGCTGAGGTGGCCAGCTGAGCT 953

Db 852 TTTAAAGGGAAGGGGAGGTGTGTGGAACCTGTAGGGGCCACAGTCCACCT 911

Qy 954 GTACATCTCGGCCCAAGCCGACCACTTT-ATTGGCACTGTGTCTCTCTTCACTG--- 1009

Db 912 GTAAATTTCTCGCAAGCCGAAACCTTTAAATTGGGAACTGTGGCCCCCTCTTCCGGGCG 971

Qy 1010 CTTTGTGTAAGCGGAGCGGACCTCCAGGGGAGGCGCTCTTCTTT 1055

Db 972 CCTTTGGAAGGGAACCGGACCTCTCAGGGAGGGCGGCTTT 1017

RESULT 6  
BX345989 841 bp mRNA linear EST 23-APR-2004  
LOCUS  
BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
DEFINITION  
BX345989 Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.  
VERSION  
BX345989.2 GI:46549776  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 841)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
COMMENT  
On May 5, 2003 this sequence version replaced gi:30377055.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: [secre@genoscope.cns.fr](mailto:secre@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
9980.f  
For more information about this cluster, see  
<http://www.genoscope.cns.fr/cdna?c=CS0AJ008DC06QP1&c=9980.f>.  
Location/Qualifiers  
1..841  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DJ008YF12"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/cell\_line="JURKAT"  
/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT  
10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.  
Library was normalized."

FEATURES  
source

Query Match 65.3%; Score 718.4; DB 5; Length 841;  
Best Local Similarity 93.5%; Pred. No. 1.1e-186;  
Matches 730; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 ATGCCCGGGCTCCTGGACCCGGCGGTTCACCTGCTTACTGCGCCCTGTCATGGGGCGC 60

Db 109 ATGCTGCGGGCTCCTGGACCCGGCGGTTCACCTGCTTACTGCGCCCTGTCATGGGGCGC 168

```
QY 61 TTGGGAACAGCCGATCACTTCTGGGCTCTGGCAATTTGCAAGTCTGTAACCGT 120
Db 169 TTGGGAACAGCCGATCACTTCTGGGCTCTGGCAATTTGCAAGTCTGTAACCGT 228
QY 121 ACCTTGGCTGCTCCCTTGGATTGAGTACCAAGTCAAGACCTCTTTACCAACCTC 180
Db 229 ACCTTGGCTGCTCCCTTGGATTGAGTACCAAGTCAAGACCTCTTTACCAACCTC 288
QY 181 CATGTGCTCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGCTTACCATCGGTCATC 240
Db 289 CATGTGCTCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGCTTACCATCGGTCATC 348
QY 241 AGCTTGGAGGATTTTCATGAGAGCTGGCACCACCCAGTCCGCCCTGAGAAGCGGGTG 300
Db 349 AGCTTGGAGGATTTTCATGAGAGCTGGCACCACCCAGTCCGCCCTGAGAAGCGGGTG 408
QY 301 GCATACCTGTTTGGAGTGGCAGCCAGCGAAGCCAGATAGAAAGCGTGCCTCATGAAG 360
Db 409 GCATACCTGTTTGGAGTGGCAGCCAGCGAAGCCAGATAGAAAGCGTGCCTCATGAAG 467
QY 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAGTCCGAG 420
Db 468 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAGTCCGAG 527
QY 421 CTTTTTACAGGCAATTCCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 480
Db 528 CTTTTTACAGGCAATTCCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 587
QY 481 CCAAGGAACATCCGGTGTGCTTCCCTGCGAGAGCCCGAGCGATTTCCCGTCTCTAGAA 540
Db 588 CCAAGGAACATCCGGTGTGCTTCCCTGCGAGAGCCCGAGCGATTTCCCGTCTCTAGAG 647
QY 541 GAACACAGCCCACTACAGAGTACATGATGTGTCAGAGAACATGATGAGAACGCGAGAG 600
Db 648 GAACACAGCCCACTACAGAGTACATGATGTGTCAGAGAACATGATGAGAACGCGAGAG 707
QY 601 GCCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTTGGCTCT 660
Db 708 GCCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTTGGCTCT 767
QY 661 GACTGGAAGACGCTGTGCCATGCTGAAGACGGGACTGCAAGGCTCGCATTCATGCCC 720
Db 768 GACTGGAAGACGCTGTGCCATGCTGAAGACGGGACTGCAAGGCTCGCATTCATGCCC 827
QY 721 TCTCCGAGTGTGT 734
Db 828 TCTCCGAGTGTGT 841
```

```
RESULT 7
LOCUS AL135434
DEFINITION DKEZp76200115_r1 762 (synonym: hmcl2) Homo sapiens cDNA clone
ACCESSION AL135434
VERSION DKEZp76200115_5', mRNA sequence.
KEYWORDS AL135434.1 GI:6603621
SOURCE EST:
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 719)
AUTHORS Anorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Anorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
```

```
German Genome Project.
No sl sequence available.
This clone (DKFZp76200115) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..719
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp76200115"
/tissue_type="melanoma (MeWo cell line)"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="762 (synonym: hmcl2)"
/note="vector: pSport1; Site_1: NotI; Site_2: SalI"
```

## ORIGIN

```
Query Match 65.2%; Score 717.4; DB 1; Length 719;
Best Local Similarity 99.9%; Pred. No. 1.9e-186;
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 360 GGAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGTGAGTTTCAACAGTCCGA 419
Db 1 GGAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGTGAGTTTCAACAGTCCGA 60
QY 420 GCTTTTACAGGCAATTCCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTC 479
Db 61 GCTTTTACAGGCAATTCCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTC 120
QY 480 TCCAAGGAACATCCGGTGTGCTTCCCTGCGAGAGCCCGAGCGATTTCCCGTCTCTAGA 539
Db 121 TCCAAGGAACATCCGGTGTGCTTCCCTGCGAGAGCCCGAGCGATTTCCCGTCTCTAGA 180
QY 540 AGAACACAGGCACTACAGAGTACATGATGTGTCAGAGAACATGATGAGAACGCGAGAG 599
Db 181 GGAACACAGGCACTACAGAGTACATGATGTGTCAGAGAACATGATGAGAACGCGAGAG 240
QY 600 GGGCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTTGGCTC 659
Db 241 GGGCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTTGGCTC 300
QY 660 TGACTGGAAGAACCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719
Db 301 TGACTGGAAGAACCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 720 CTCTCCGAGTGTGTGGGCTACAGCCGAGACAGCGGCCCGCCCTCAGATGACTATGTG 779
Db 361 CTCTCCGAGTGTGTGGGCTACAGCCGAGACAGCGGCCCGCCCTCAGATGACTATGTG 420
QY 780 CTTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGC 839
Db 421 CTTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGC 480
QY 840 CCAGTCTGCTTACCTTGTCTACTGATTCGAGAGTTATGTGCTGCTGCTGCTGCTGCTGCT 899
Db 481 CCAGTCTGCTTACCTTGTCTACTGATTCGAGAGTTATGTGCTGCTGCTGCTGCTGCTGCT 540
QY 900 CAAAGGAAGGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 959
Db 541 CAAAGGAAGGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 600
QY 960 CTTGGCCAGCCCAACCATTTTATTTGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1019
Db 601 CTTGGCCCAAGCCCAACCATTTTATTTGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
QY 1020 GCGGGAGCGGACCTCCAGGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1078
Db 661 GCGGGAGCGGACCTCCAGGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719
```

RESULT 8  
BI524031



```
Query Match      64.2%; Score 706.6; DB 5; Length 898;
Best Local Similarity 97.8%; Pred. No. 2e-183;
Matches 744; Conservative 4; Mismatches 10; Indels 3; Gaps 3;

QY 1 ATGCCCGGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCTCCCTGCGATGGGGCG 60
DB 139 ATGCCCGGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCTCCCTGCGATGGGGCG 198

QY 61 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120
DB 199 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 258

QY 121 ACCTTGGGCTGCTCCCTCTTGGATTTGAGTACAGATCAAGACCTCTCTTTTCAACCACTC 180
DB 259 ACCTTGGGCTGCTCCCTCTTGGATTTGAGTACAGATCAAGACCTCTCTTTTCAACCACTC 318

QY 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
DB 319 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 378

QY 241 AGCTTGGAGGATTTATGAGAGAGTGGCAACCCACTGGCCCTCGAGAAGCGGGTG 300
DB 379 AGCTTGGAGGATTTATGAGAGAGTGGCAACCCACTGGCCCTCGAGAAGCGGGTG 438

QY 301 GCATATCTGTTTGGGTGGCAGCCAGCGAAGCCAGATTAAGAAGACGTGCCCATGAAG 360
DB 439 GCATATCTGTTTGGGTGGCAGCCAGCGAAGCCAGATTAAGAAGACGTGCCCATGAAG 497

QY 361 GAAGGAACCCCTTTGGCCATCTTGGGATCTGGGATCAAGTATGATGTTTCAAGTCGGAG 420
DB 498 GAAGGAACCCCTTTGGCCATCTTGGGATCTGGGATCAAGTATGATGTTTCAAGTCGGAG 557

QY 421 CTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 480
DB 558 CTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 617

QY 481 CCAAAGGAACATCCGGTGTGCTCCCTGCGAGGAGCCAGCCAGTTCCTCCCTCTAGAG 540
DB 618 CCAAAGGAACATCCGGTGTGCTCCCTGCGAGGAGCCAGCCAGTTCCTCCCTCTAGAG 677

QY 541 GAACACAGGCACTACAGAGTACATGTATGGTCAGAG- AATGGTGAAGACGGGAGA 599
DB 678 GAACACAGGCACTACAGAGTACATGTATGGTCAGAGCAAAATGGTGAAGACGGGAGA 737

QY 600 GSCCCAGATTCATGCCACCTTGTCCGCTGAGGAGCCAGCCAGTTCATCTGCGCATGGCTC 659
DB 738 GSCCCAGATTCATGCCACCTTGTCCGCTGAGGAGCCAGCCAGTTCATCTGCGCATGGCTC 797

QY 660 TGACTGGAAGAACGCTGTGCAATGCTGAAGGACGGGACTGAGGCTCGCACTTCAATGGC 719
DB 798 TGACTGGAAGAACGCTGTGCAATGCTGAAGGACGGGACTGAGGCTCGCACTTCAATGGC 857

QY 720 CT-CTCCGAGTGTGGGCTACAGCGGAGCAGCGGCC 759
DB 858 CTCTCTCCGAGTGTGGGCTACAGCGGAGCAGCGGCC 898

RESULT 10
CO774833 LOCUS
DEFINITION ILLUMIGEN MCQ 51950 Katze MNLV Macaca nemestrina cDNA clone
            IBUW:29950 57 similar to Bases 7 to 852 highly similar to human
            POFU1 (Hs.178292), mRNA sequence.
ACCESSION CO774833
VERSION CO774833.1 GI:50976096
KEYWORDS EST.
SOURCE Macaca nemestrina (pig-tailed macaque)
ORGANISM Macaca nemestrina
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
            Cercopitheciinae; Macaca.
REFERENCE 1 (bases 1 to 942)
```

AUTHORS  
TITLE  
JOURNAL  
COMMENT

Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magness, C.L.  
Large-scale Rhesus Macaque cDNA Sequencing  
Unpublished (2003)  
Contact: C. Magness  
Illumigen Biosciences Inc.  
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA  
Tel: 2063780400  
Fax: 2063780408  
Email: cmagness@illumigen.com  
Sequenced on 2004.07.21. 611 020 bases. Library Preparation: Prof.  
Michael Katze Lab at University of Washington DNA Sequencing:  
Illumigen Biosciences Inc. For further information, see  
http://www.macaque.org

## PCR Primers

FORWARD: CCCTCACTAAAGGGAACAAA

BACKWARD: CACTATAGGCGAATTGGGTA

Insert Length: 942 Std Error: 0.00

Plate: CL000531 row: E column: 09

Seq primer: CCCTCACTAAAGGGAACAAA

POLYA=No.

## Location/Qualifiers

1. 942

/organism="Macaca nemestrina"

/mol\_type="mRNA"

/strain="Indian"

/db\_xref="taxon:9545"

/clone="IBUW:29960"

/sex="male"

/lab\_host="Electronax DH10B"

/clone\_lib="Katze MNLV"

/note="Organ: liver; Vector: pDONR 222; Site 1: BsrG I;

Site 2: BsrG I; Created from Cloneminer cDNA Library

Construction kit (catalog #18249-029)"

## ORIGIN

Query Match 61.2%; Score 673.2; DB 7; Length 942;  
Best Local Similarity 94.1%; Pred. No. 3.3e-174;  
Matches 743; Conservative 0; Mismatches 43; Indels 4; Gaps 4;

QY 1 ATGCCCGGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCTCCCTGCGATGGGGCG 60  
DB 95 ATGCCCGGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCTCCCTGCGATGGGGCG 154

QY 61 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
DB 155 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGC 214

QY 121 ACCTTGGGCTGCTCCCTCTTGGATTTGAGTACAGATCAAGACCTCTCTTTTCAACCACTC 180  
DB 215 ACCTTGGGCTGCTCCCTCTTGGATTTGAGTACAGATCAAGACCTCTCTTTTCAACCACTC 274

QY 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
DB 275 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 334

QY 241 AGCTTGGAGGATTTTCAAGGAGAGCTGGCAACCCACTGCGCCCTGAGAAGCGGGTG 300  
DB 335 AGCTTGGAGGATTTTCAAGGAGAGCTGGCAACCCACTGCGCCCTGAGAAGCGGGTG 394

QY 301 GCATATCTGTTTGGGTGGCAGCCAGCGAAGCCAGATTAAGAAGACGTGCCCATGAAG 360  
DB 395 GCATATCTGTTTGGGTGGCAGCCAGCGAAGCCAGATTAAGAAGACGTGCCCATGAAG 454

QY 361 GAAGGAACCCCTTTGGCCATCTTGGGATCTGGGATCAGTTTCAATGAGTTTCAACAGTCGGAG 420  
DB 455 GAAGGAACCCCTTTGGCCATCTTGGGATCTGGGATCAGTTTCAATGAGTTTCAACAGTCGGAG 514

QY 421 CTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 480  
DB 515 CTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 574

QY 481 CCAAAGGAACATCCGGTGTGCTCCCTGCGAGGAGCCAGCCAGTTCCTCCCTCTAGAA 540



National Institute of Agrobiological Sciences  
2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan.  
Tel: +81-29-838-8627  
Fax: +81-29-838-8627

Email: huenishia@affrc.go.jp

EST project with full-length enriched cDNA libraries carried out in  
Animal Genome Research Program (Japan) by National Institute of  
Agrobiological Sciences and STAFF-Institute  
Single pass sequencing of clones derived from oligo-capped cDNA  
library

Vector sequences were eliminated by RepeatMasker version 2002/07/13  
and crossmatch version 0.990319  
Low quality bases were trimmed based on the quality values.

## FEATURES

source

1. 801  
Location/Qualifiers  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/clone="OVR010034G09"  
/tissue type="ovary"  
/dev stage="adult"  
/clone\_lib="full-length enriched swine cDNA library, adult  
ovary"

## ORIGIN

Query Match 60.0%; Score 660.2; DB 5; Length 801;  
Best Local Similarity 89.0%; Pred. No. 1.2e-170;  
Matches 713; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 194 AGAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCTATCAGCTTGGAGGATT 253  
DB 1 AAAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCTATCAGCTTGGAGGACT 60  
QY 254 TCATGGAGAAGTGGCACCACCCACTGGCCCTGAGAGGGGTGCATCTGCTTTG 313  
DB 61 TCATGGAGAAGTGGCACCACCCACTGGCCCTGAGAGGGGTGCATCTGCTTTG 120  
QY 314 AGGTGGCAGCCAGCGAAGCCAGATAGAGAGCTGCCCATGAAGGAAGAAACCCCT 373  
DB 121 AGGTGGCAGCAGCGAAGCCAGATAGAGAGCATGCCCCATGAAGGAAGAAACCCCT 180  
QY 374 TTGGCCCATTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTACAGGCA 433  
DB 181 TTGGCCCATTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCAGCACTTTTTCAGGCA 240  
QY 434 TTTCTTCAGTGTCTTCTACAGAGAAATATGGAGCCAGAGATTTTCTCAAGGAACATC 493  
DB 241 TTTCTTCAGTGTCTTCTACAGAGAACCAAGAGATTTTCTCAAGGAACATC 300  
QY 494 CGGTGTTGCCCTGCCAGGAGCCCGAGCCAGTTCCTCCCGTCTAGAGAAACACAGGCCAC 553  
DB 301 CAGTGTGTTGCCCTGCCAGGAGCCCGGTCTAGTTCCTCCCGTCTAGAGAAACATAGGCCAC 360  
QY 554 TACAGAAGTACATGTATGTGTCAGACGAAATGGTGAAGACGGGAGGCCAGATTCATG 613  
DB 361 TTCAGAATACATGTGTGGTCAGACGAGATGGTGAAGACGGGAGGCCAGATCCGTG 420  
QY 614 CCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGGCGATTTGCTGTGACCTGGAAGAACG 673  
DB 421 CCCACCTTATCCGGCCCTATGTGGGCATTCATCTGGCGATTTGCTGTGACCTGGAAGAACG 480  
QY 674 CTTGTGCCATCTGAAGACGGGACTGCAGGCTCGACCTTCATGCTCTCCGAGTGTG 733  
DB 481 CGTGGCCATCTGAAGACGGGACTGCAGGCTCGACCTTCATGCTCTCCGAGTGTG 540  
QY 734 TGGGCTACAGCCGAGCAGCGGCCCTCCATGATGATGTGCTGCTGACCTGA 793  
DB 541 TGGGTTACAGCCGACACACCGGCCCTCCATGATGATGTGCTCTCCGAGTGTG 600  
QY 794 AGGAGATCCAGAGGCTGTGAAGTCTTGGGTGAGGTGCTGGATGCTCCAGTGTGCTACG 853  
DB 601 AGGAGATCAGCGGCCCTGGAAGTCTTGGGTGAGGTGCTGGCGCTGAGCGGCCCGCTCTACA 660

QY 854 TTGCTACTGATTCCGAGAGTTTATGTGCTGAGCTCCACAGCTCTTCAAAGGGAAGGTGA 913  
DB 661 TCGCCACCGACTCTGAGAGTTTACTTCCGAGAGATCCAGCAGCTCTTCAAAGGGAAGGTGA 720  
QY 914 AGGTGGTGAAGCTGAAGCTGAGGTGGCCAGGTGAGCTGTACATCTCCGCCCAAGCCG 973  
DB 721 AGGTGGTGAAGCTGAAGCTGAGGTGGCCAGGTGAGCTGTACATCTTGGCCAGGTG 780  
QY 974 ACCACTTTATTTGGCAACTGTG 994  
DB 781 ACCACTTTATTTGGCAACTGTG 801

## RESULT 13

BO964179

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

BO964179

AGENCY

IMAGE:6509783 5', mRNA sequence.

BO964179

EST.

GI:22379657

Mus musculus

(house mouse)

|    |     |   |                                   |     |
|----|-----|---|-----------------------------------|-----|
| Db | 328 | TGTTGCTCTACCAAAAGTACTCTAAACTGGAGAGCTCTC | CAAGCCTTACCATCGGTTGTGCAG          | 388 |
| Qy | 243 | CTTGGAGGATTCATGGAGAGCTGGCACCCACCCACCT   | CTGGCCCCCTGGAGAGCGGGTGGC          | 302 |
| Db | 388 | CCTGGAGGACTTCATGGAAATCTTGGACACCTCCCACT  | TGGCCCCCTGGAGAGCGAGTGGC           | 447 |
| Qy | 303 | ATACTGCTTTTGAGGTGGCAGCCACGACGAAAGCC     | CACAGATAAGAAAGACGTGCCCATGAAGA     | 362 |
| Db | 448 | ATACTGCTTTTGAGGTGGCAGCCACGACGAAAGTCT    | GATAAGAAAGACATGTCCCATGAAGA        | 507 |
| Qy | 363 | AGGAAACCCCTTTGGGCCAATCTCGGATCAGTTTCA    | TGTGAGTTTCAACAAGTCGGAGCT          | 422 |
| Db | 508 | AGGAAATCCTTTTGGGCCAATCTGGACCAAGTTTCA    | TGTGAGTTTCAATAGTCAGACT            | 567 |
| Qy | 423 | TTTTTACAGGCAATTCCTTCAGTGCTTCTACAGAA     | CAATGGAGCCAGAGATTTCTCC            | 482 |
| Db | 568 | GTTTCACAGGCAATTCCTTCAGGGCTCTCTACAAG     | AACATGGACCCAGAGATTTCTCC           | 627 |
| Qy | 483 | AAAGAAACATCGGTGCTTTGGCCCTGCCAGAGCC      | CCAGCCAGTTTCCCGTCTTAGAGA          | 542 |
| Db | 628 | AAAAGAGACATCTGTGTCTCGCACTGCTTGGG        | GGCCCCACGACAGTTTCCGTCTCTGGAGGA    | 687 |
| Qy | 543 | ACACAGGCCACTACAGAAGTACATGTTATGGT        | CAGACGAAATGGTGAAGACGGGAGAGCC      | 602 |
| Db | 688 | ACACAGGGAGCTCCAGAAGTACATGTTGTGGT        | CAGATGAGATGGTGAAGACGGGAGAGCC      | 747 |
| Qy | 603 | CCGATTCATGCCCACTTGTCGGGCCCTATGTGG       | CACTTGTGGCAATTCATCTGGCACTGGCTCTGA | 662 |
| Db | 748 | CCTGATCAGTGCCCACTCTGTCGGGCCCTATGT       | GGGCATTCATCTGGGCATTTGGCTCCGA      | 807 |
| Qy | 663 | CTGGAAGAACGCTGTGCCATGCTGGAAGGACGG       | GAATGCGAGCTCGCACTTCATATGCGCTTC    | 722 |
| Db | 808 | CTGGAAGAAATGCTGTGCCATGCTGGAGGATGAA      | CTGCAGGGTCACACTTCATATGCGCTTC      | 867 |
| Qy | 723 | TCC-GCAGTGTGGGCTACAGCCGACGACACAG        | CGCCCCCTCCACGATGACTATGTGCC        | 781 |
| Db | 868 | CCCTTCNNTGTGTGGGCTATAGCGCAGCAGCG        | ACCCCTCTCAC-ATGACCATGTGCC         | 926 |
| Qy | 782 | TGCCTGACCTGAAGAGAGATCCACAGGGGCTGT       | GAAGCTCTGGGTGAG                   | 827 |
| Db | 927 | TCCCTGACCTGAAGAAATACAGCGGCTGGGAA        | CGCTTTGGGGTG                      | 976 |

RESULT 14  
BI757540  
LOCUS  
BI757540 874 bp mRNA linear EST 25-SEP-2001  
603029693F1 NIH\_MGC\_114 Homo sapiens CDNA clone IMAGE:519862 5',  
DEFINITION  
mRNA sequence.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 874)  
1 (bases 1 to 874)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

FEATURES  
SOURCE

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5199862"
/lab_host="DH10B"
/clone_lib="NTH MGC 114"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NTH MGC library."

```

## ORIGIN

| Query Match           | 57.9%;       | Score 636.4;   | DB 4;     | Length 874; |
|-----------------------|--------------|--|-----------|-------------|
| Best Local Similarity | 96.1%;       | Prod. No. 4.5e-164;  |           |             |
| Matches 717;          | Conservative | 0; Mismatches 21;  | Indels 8; | Gaps 6;     |
| QY                    | 1            | ATGCCCGCGGGCTCTCTGGGACCCGGCGGGTTACTGTCTACTGCCCCCTGCATGGGGCGC     | 60        |             |
| DB                    | 109          | ATGCCCGCGGGCTCTGGGACCCGGCGGGTTACTGTCTACTGCCCCCTGCATGGGGCGC       | 168       |             |
| QY                    | 61           | TTTGGGAAACCGCGCGATCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACG-G      | 119       |             |
| DB                    | 169          | TTTGGGAAACCGCGCGATCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCTG      | 228       |             |
| QY                    | 120          | TACTTTGGCTGTCCCTCTCTTTGGATTTGAGTACAGATCAACAAGCTCTCTTTCACAACCT    | 179       |             |
| DB                    | 229          | TACTTTGGCTGTCCCTCTCTTTGGATTTGAGTACAGATCAACAAGCTCTCTTTCACAACCT    | 288       |             |
| QY                    | 180          | CCATGTGTCTTACCAGAAGTACTTCAAGCTGGAGCGCCCTCCAGGCTTACCATCGGGTCAT    | 239       |             |
| DB                    | 289          | CCATGTGTCTTACCAGAAGTACTTCAAGCTGGAGCGCCCTCCAGGCTTACCATCGGGTCAT    | 348       |             |
| QY                    | 240          | CAGCTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCACTGGCCCCCTTGAGAAGCGGGT    | 299       |             |
| DB                    | 349          | CAGCTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCACTGGCCCCCTTGAGAAGCGGGT    | 408       |             |
| QY                    | 300          | GGCATACTGCTTTGAGTGGGAGCCAGCCAGGAGCCAGATAAGAGACGTGCCCATGAA        | 359       |             |
| DB                    | 409          | GGCATACTGCTTTGAGTGGGAGCCAGGAGGAGCCAGATAAGAGACGTGCCCATGAA         | 468       |             |
| QY                    | 360          | GGAGGAAACCCCTTTGGCCCATCTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA     | 419       |             |
| DB                    | 469          | GGAGGAAACCCCTTTGGCCCATCTCTGGGATCAG-TTCAATGTGAGTTTCAACAAGTCGGA    | 527       |             |
| QY                    | 420          | GCTTTTTTACAGGCAATTCCTTTCAAGTGCTTCTTACAGAGAACAAATGGAGCCAGAGATTTTC | 479       |             |
| DB                    | 528          | GCTTTTTTACAGGCAATTCCTTTCAAGTGCTTCTTACAGAGAACAAATGGAGCCAGAGATTTTC | 587       |             |
| QY                    | 480          | TCCAAAGGAACATCCGGTGCTTGCCCTGCGAGAGCCCC-AGCCGAGTTCCCGTCCTAG       | 538       |             |
| DB                    | 588          | TCCAAAGGAACATCCGGTGCTTGCCCTGCGAGAGCCCCAAGCCGAGTTTCCCGTCCTAG      | 647       |             |
| QY                    | 539          | AAGAACACAGGGCCACTACAGAAGTACATGGTATGCTCAGAC--GAAATGGTGAAGACGGG    | 596       |             |
| DB                    | 648          | AAGAACACAGGGCCACTACAGAAGTACATGGTATGCTCAGACCCGACCAATGGTGAAGACGGG  | 707       |             |
| QY                    | 597          | AGAGGCCAGATTTCAATGCCACCTTTGTTCGGCCCTTATGTGGGCATTTCATCTGCGATTGG   | 656       |             |
| DB                    | 708          | AGAGGCCAGATTTCAATGCCACCTTTGTTCGGCCCTTATGTGGGCATTTCATCTGCGATTGG   | 767       |             |
| QY                    | 657          | CT-CTGACTGGAGAACGCCCTGTGCCATGCTGA--AGACGGGACTGCAAGGCTCGCACTT     | 713       |             |
| DB                    | 768          | CTCTGACTGGAGGAACGCCCTGTGCCATGCTGACCCGAGCTGGGCTTGCAGGCTTCCCTTC    | 827       |             |
| QY                    | 714          | CATGGGCTCTCCGCAAGTGTGGGCT  | 739       |             |
| DB                    | 828          | ATGGCTCTCTCCGCAAGTGTGGGCT  | 853       |             |



## RESULT 15

BI410783  
LOCUS 602963556F1 NCI\_CGAP\_Lu33 947 bp mRNA linear EST 14-AUG-2001  
DEFINITION mRNA sequence.  
ACCESSION BI410783  
VERSION BI410783.1 GI:15171706  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 947)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM11290 row: o column: 13  
High quality sequence start: 21  
High quality sequence stop: 919.

## FEATURES

## source

1..947  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5119332"  
/tissue\_type="pooled lung tumors"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NCI CGAP Lu33"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site1: NotI; Site2: EcoRI; 1st  
strand cDNA was prepared from mRNA obtained from pooled  
lung tumors with a Not I - oligo(dT) primer [5].  
TGTACCAATCTGAAGTGGAGCGCGCTCTGTTTCTTTTCTTTT 3'.  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 57.3%; Score 630.8; DB 4; Length 947;  
Best Local Similarity 86.9%; Pred. No. 1.6e-162;  
Matches 763; Conservative 0; Mismatches 107; Indels 8; Gaps 6;  
QY 3 GCCCGCGGCTCTGGGACCGCGGTACTGCTCTACTGCCCTGCATGGGCGCTT 62  
DB 70 GTCAGGGGCTCTGGGACCTGGCGGTACTGCTCTACTGCTCCTGCTATGGGCGCTT 129  
QY 63 TGGGAACAGGCGCATCATCTTTTGGGCTCTCTGGCAATTCGCAAGCTGTAAACCGTAC 122  
DB 130 TGGGAACAGGCGTATCATCTTTGGGCTCTCTGGCAATTCGCAAGCTGTAAACCGCAC 189  
QY 123 CTTGGCTCTCCTCTCTGGATGAGTACCAGCATCAGAGCTCTCTTTCACCACTCCA 182  
DB 190 CTTGGCTGTACTCTCATGATTTGAATACCAACAGGCTCTCTTTCACCACTCCA 249  
QY 183 TGTGTCTTACCAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCATCAG 242  
DB 250 TGTGTCTTACCAGAGTACTTCAAACTGGAGCTCTCTCAAGCTTACCATGGGTTGTAC 309  
QY 243 CTTGGAGGATTCATGGAGAGCTGGCACCCACCTCTGGCCCTCGAGAGCGGTGGC 302

DB 310 CTTGGAGGACTTTCATGAAATCTGGCACCCCTCCCACTGGCCCTCGAGAGCGAGTGGC 369  
QY 303 ATACTGCTTTGAGTGGCAGCCGCGAGCCAGATGAAGAGCTGCCCATGAAGA 362  
DB 370 ATACTGCTTTGAGTGGCAGCCGCGAGTCTCTGATGAAGAGCATGTCCTCATGAAGA 429  
QY 363 AGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCT 422  
DB 430 AGGAAATCTTTTGGCCATTTCTGGACCAAGTTTCATGTGAGTTTCAATAGTCAGAACT 489  
QY 423 TTTTACAGGCATTTCTTTCACTGCTTCTCTACAGAGAACTAGGAGCCAGAGATTTCTCC 482  
DB 490 GTTCACAGGCAATTTCTTTCAAGGCTCTCTCAAAAGAACTAGGAGCCAGAGATTTCTCTGC 549  
QY 483 AAAGGAACATCCGCTGCTTGGCCCTGCCAGGAGCCCGCCAGTTCCTCCGT--CCTAGAG 541  
DB 550 AAAAGAGCATCTGTGCTCGCACTGCTGGGGCCCCAGCACAGTTTCCCTCTGGAGG 609  
QY 542 AACACAGGCCACTACAGAAATGATGCTGATGCTCAGACGAAATGGTGAAGACGGGAGAGG 601  
DB 610 AACACAGGGAGCTCCAGAAATGATGCTGCTGCTGATGATGATGATGATGATGATGATGATG 669  
QY 602 CCCAGATTCATGCCCCACTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCAATTTGGCTCTG 661  
DB 670 CCTGATCAGTGGCCCACTGCTCGGGCCCTATGTGGGCATTTCATCTGGCA--TGGCTCG 728  
QY 662 ACTGGAAGACGCTGTGCCATGCTGAAGGAGGGAGCTGAGGCTCGGCTCATGGCT 721  
DB 729 ACTGGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 788  
QY 722 CTCCGCACTGTGGGCTTACAGCCGACAGCGGCCCTCCATGATGATGATGATGATGATGATGATG 781  
DB 789 CCCCTCAGTGTGGGCTATAGCCGACGACCAAGACCCCTC--TCACATGACCATGTGCC 846  
QY 782 TGCCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTTGGGTGAGGTGCTGATGCC 841  
DB 847 TCCCTGACCTGAAGGAAAT--CAGCGGGCTGTGAGGCTTTGGGTGAGAGCATTTG--ATGCC 903  
QY 842 AGTCGGTCTAGTTGCTACTGATTTCCGAGAGTTATGTG 879  
DB 904 AATCGGTCTACATGCG--ACAGACTCTGAGAGCTACGTG 940

Search completed: November 9, 2005, 10:38:07

Job time : 3990.87 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 08:00:04 ; Search time 984.698 Seconds  
(without alignments)  
9238.165 Million cell updates/sec

Title: US-09-774-954-16  
Perfect score: 1100  
Sequence: 1 ATGCCCGGGGCTCTCTGGGA.....CTGCGGAGGAGTCTTCATT 1100

Scoring table: IDENTITY NUC.  
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/FCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/FCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*  
23: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
24: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*  
26: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
27: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
28: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description       |
|------------|--------|-------------|--------|----|-------------------|
| 1          | 1100   | 100.0       | 1100   | 11 | US-09-774-954-16  |
| 2          | 1100   | 100.0       | 1514   | 11 | US-09-774-954-1   |
| 3          | 1098.4 | 99.9        | 1300   | 11 | US-09-774-954-4   |
| 4          | 1098.4 | 99.9        | 11284  | 11 | US-09-774-954-5   |
| 5          | 1096.8 | 99.7        | 5218   | 16 | US-10-301-822-168 |

|    |        |      |         |    |                       |                    |
|----|--------|------|---------|----|-----------------------|--------------------|
| 6  | 1096.8 | 99.7 | 5218    | 22 | US-10-956-157-4752    | Sequence 4752, Ap  |
| 7  | 1033.4 | 93.9 | 5009    | 11 | US-09-774-954-7       | Sequence 7, Appli  |
| 8  | 471.4  | 42.9 | 1400    | 22 | US-10-956-157-9986    | Sequence 9986, Ap  |
| 9  | 471.4  | 42.9 | 1506    | 22 | US-10-956-157-4751    | Sequence 4751, Ap  |
| 10 | 251.2  | 22.8 | 10331   | 10 | US-09-764-891-6032    | Sequence 6032, Ap  |
| 11 | 233    | 21.2 | 479     | 10 | US-09-918-995-19437   | Sequence 19437, A  |
| 12 | 221.8  | 20.2 | 477     | 10 | US-09-918-995-22744   | Sequence 22744, A  |
| 13 | 166.4  | 15.1 | 1209    | 26 | US-11-097-143-18161   | Sequence 18161, A  |
| 14 | 150.4  | 13.7 | 3264    | 26 | US-11-097-143-18160   | Sequence 18160, A  |
| 15 | 114.6  | 10.4 | 402     | 9  | US-09-783-590-11501   | Sequence 11501, A  |
| 16 | 108.6  | 9.9  | 3793    | 26 | US-11-097-143-18082   | Sequence 18082, A  |
| 17 | 71.8   | 6.5  | 3567    | 26 | US-11-097-143-18169   | Sequence 18169, A  |
| 18 | 60     | 5.5  | 60      | 10 | US-09-908-975-17287   | Sequence 17287, A  |
| 19 | 56     | 5.1  | 760     | 24 | US-10-820-474A-236    | Sequence 236, App  |
| 20 | 55.6   | 5.1  | 521     | 13 | US-09-925-065A-60244  | Sequence 60244, A  |
| 21 | 42.4   | 3.9  | 1438    | 18 | US-10-282-122A-19431  | Sequence 19431, A  |
| 22 | 41.2   | 3.7  | 114793  | 16 | US-10-148-806-3       | Sequence 3, Appli  |
| 23 | 41.2   | 3.7  | 114793  | 24 | US-10-859-792-3       | Sequence 3, Appli  |
| 24 | 40.4   | 3.7  | 1212    | 21 | US-10-425-115-114097  | Sequence 114097, A |
| 25 | 40.4   | 3.7  | 5932    | 16 | US-10-205-823-312     | Sequence 312, App  |
| 26 | 40.4   | 3.7  | 5932    | 26 | US-11-051-454-312     | Sequence 312, App  |
| 27 | 40     | 3.6  | 40      | 11 | US-09-774-954-10      | Sequence 10, Appl  |
| 28 | 40     | 3.6  | 40      | 11 | US-09-774-954-12      | Sequence 12, Appl  |
| 29 | 40     | 3.6  | 2032    | 20 | US-10-437-963-8098    | Sequence 8098, Ap  |
| 30 | 39.2   | 3.6  | 618     | 13 | US-09-925-065A-571171 | Sequence 571171, A |
| 31 | 39     | 3.5  | 372     | 18 | US-10-282-122A-31476  | Sequence 31476, A  |
| 32 | 38.6   | 3.5  | 407     | 9  | US-09-960-352-11732   | Sequence 11732, A  |
| 33 | 38.6   | 3.5  | 2358    | 10 | US-09-919-039-190     | Sequence 190, App  |
| 34 | 38.6   | 3.5  | 2384    | 22 | US-10-278-698-123     | Sequence 123, App  |
| 35 | 38.6   | 3.5  | 2384    | 22 | US-10-278-698-638     | Sequence 638, App  |
| 36 | 38.6   | 3.5  | 2670    | 17 | US-10-247-671-64      | Sequence 64, Appl  |
| 37 | 38.4   | 3.5  | 532     | 10 | US-09-918-995-13517   | Sequence 13517, A  |
| 38 | 38.4   | 3.5  | 591     | 16 | US-10-156-761-3550    | Sequence 3550, Ap  |
| 39 | 38.4   | 3.5  | 6227    | 19 | US-10-240-425-364     | Sequence 364, App  |
| 40 | 38.4   | 3.5  | 6227    | 22 | US-10-278-698-100     | Sequence 100, App  |
| 41 | 38.4   | 3.5  | 6227    | 22 | US-10-278-698-615     | Sequence 615, App  |
| 42 | 38.4   | 3.5  | 6256    | 22 | US-10-278-698-101     | Sequence 101, App  |
| 43 | 38.4   | 3.5  | 6256    | 22 | US-10-278-698-616     | Sequence 616, App  |
| 44 | 38.4   | 3.5  | 9025608 | 16 | US-10-156-761-1       | Sequence 1, Appli  |
| 45 | 37.8   | 3.4  | 1234    | 18 | US-10-369-493-39272   | Sequence 39272, A  |

ALIGNMENTS

RESULT 1  
US-09-774-954-16  
; Sequence 16, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Pucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498



Qy 1 ATGCCCCGGGGCTCTGGGACCCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Db 1 ATGCCCCGGGGCTCTGGGACCCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Qy 61 TTTGGGAACAGGCGCGATCACTTCTTGGGCTCTTGGGATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACAGGCGCGATCACTTCTTGGGCTCTTGGGATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGGCTGTCCTTCTTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 180  
Db 121 ACCTTGGGCTGTCCTTCTTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 180  
Qy 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCACTC 240  
Db 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCACTC 240  
Qy 241 AGCTTGGAGGATTTTCATGGAAGCTGGCAGCCACCCACCTGCGCCCTGAGAAGCGGGT 300  
Db 241 AGCTTGGAGGATTTTCATGGAAGCTGGCAGCCACCCACCTGCGCCCTGAGAAGCGGGT 300  
Qy 301 GCATACTCTTTGAGTGGCAGCCAGGAGAGCCAGATAGAGAGCTGCCCATGAG 360  
Db 301 GCATACTCTTTGAGTGGCAGCCAGGAGAGCCAGATAGAGAGCTGCCCATGAG 360  
Qy 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAGTCGGAG 420  
Db 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAGTCGGAG 420  
Qy 421 CTTTTCAGGATTTCTTCAGTGTCTTCTACAGAGAAATGAGGACGAGATTTTCT 480  
Db 421 CTTTTCAGGATTTCTTCAGTGTCTTCTACAGAGAAATGAGGACGAGATTTTCT 480  
Qy 481 CCAAGGAACATCGGTCTGCTGCGCCAGGAGCCAGCCAGCTTCCCGTCTAGAA 540  
Db 481 CCAAGGAACATCGGTCTGCTGCGCCAGGAGCCAGCCAGCTTCCCGTCTAGAA 540  
Qy 541 GAACACAGGCCATACAGAGTACATGTATGTGTGAGAGAAATGAGGAGCGGAGAG 600  
Db 541 GAACACAGGCCATACAGAGTACATGTATGTGTGAGAGAAATGAGGAGCGGAGAG 600  
Qy 601 GCCAGATTCATGCCCACCTTGTGCGGCCCTATGTGGGATTCATCTGCGATTTGGCTCT 660  
Db 601 GCCAGATTCATGCCCACCTTGTGCGGCCCTATGTGGGATTCATCTGCGATTTGGCTCT 660  
Qy 661 GACTGGAGAGCGCTGTGCTGCTGAGGAGCGGAGCTGAGGCTGCGACTTATGGCC 720  
Db 661 GACTGGAGAGCGCTGTGCTGCTGAGGAGCGGAGCTGAGGCTGCGACTTATGGCC 720  
Qy 721 TCTCCGAGTGTGTGGGCTACAGCGGAGCAGCAGCGGCCCTCAGCATGACTATGTGC 780  
Db 721 TCTCCGAGTGTGTGGGCTACAGCGGAGCAGCAGCGGCCCTCAGCATGACTATGTGC 780  
Qy 781 CTGCTGACCTGAAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGATGCTGCC 840  
Db 781 CTGCTGACCTGAAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGATGCTGCC 840  
Qy 841 CAGTGGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTTGGCTGAGCTTCAACGCTTTC 900  
Db 841 CAGTGGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTTGGCTGAGCTTCAACGCTTTC 900  
Qy 901 AAAGGGAAGTGAAGTGTGCTGAGCTGAAGCTCAGGTTGGCCAGGCTGCTGATC 960  
Db 901 AAAGGGAAGTGAAGTGTGCTGAGCTGAAGCTCAGGTTGGCCAGGCTGCTGATC 960  
Qy 961 CTCGCGCAAGCCGACCACTTTATTGGCAATGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 961 CTCGCGCAAGCCGACCACTTTATTGGCAATGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Qy 1021 CGGAGCGGAGCTTCCAGGAGGCGCTTCTTCTTTCGCGATGAGCAGGCCCTTAAG 1080  
Db 1021 CGGAGCGGAGCTTCCAGGAGGCGCTTCTTCTTTCGCGATGAGCAGGCCCTTAAG 1080

Qy 1081 CTGCGGGACGAGTCTTGATT 1100  
Db 1081 CTGCGGGACGAGTCTTGATT 1100  
RESULT 3  
US-09-774-954-4  
; Sequence 4, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498  
; FILING DATE: 31-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1300 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-774-954-4

Query Match 99.9%; Score 1098.4; DB 11; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCCGGGCTCTGGGACCCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Db 136 ATGCCCCGGGCTCTGGGACCCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 195  
Qy 61 TTTGGGAACAGGCGCGATCACTTCTTGGGCTCTCTGGGATTTGCAAGCTGCTAAACCGT 120  
Db 196 TTTGGGAACAGGCGCGATCACTTCTTGGGCTCTCTGGGATTTGCAAGCTGCTAAACCGT 255  
Qy 121 ACCTTGGGCTGTCCTCTTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 180  
Db 256 ACCTTGGGCTGTCCTCTTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 315  
Qy 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCACTC 240  
Db 316 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCACTC 375  
Qy 241 AGCTTGGAGGATTTTCATGGAAGCTGGCAGCCACCCACCTGCGCCCTGAGAAGCGGGTG 300  
Db 376 AGCTTGGAGGATTTTCATGGAAGCTGGCAGCCACCCACCTGCGCCCTGAGAAGCGGGTG 435

```
QY 301 GCATACCTGCTTTGAGTGGACCCAGCGAAGCCAGATGAAGACGTCGCCCATGAAG 360
Db 436 GCATACCTGCTTTGAGTGGACCCAGCGAAGCCAGATGAAGACGTCGCCCATGAAG 495
QY 361 GAAGGAACACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db 496 GAAGGAACACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 555
QY 421 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAACCAATGGAGCCAGAGATTTTCT 480
Db 556 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAACCAATGGAGCCAGAGATTTTCT 615
QY 481 CCAAGGAACATCCGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 616 CCAAGGAACATCCGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 675
QY 541 GAACACAGGCACTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 676 GAACACAGGCACTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735
QY 601 GCCCAGATTTCATGCCACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 736 GCCCAGATTTCATGCCACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 795
QY 661 GACTGGAAGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 796 GACTGGAAGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 855
QY 721 TCTCGCAGTGTGTGGGCTACAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGTGC 780
Db 856 TCTCGCAGTGTGTGGGCTACAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGTGC 915
QY 781 CTGCTGACCTGAGGAGATCAGAGGCTGCTGAAGCTCTGGTGAAGTCTGGTGAAGTCTGGTGAAG 840
Db 916 CTGCTGACCTGAGGAGATCAGAGGCTGCTGAAGCTCTGGTGAAGTCTGGTGAAGTCTGGTGAAG 975
QY 841 CAGTGGTCTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 976 CAGTGGTCTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1035
QY 901 AAAGGAAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 960
Db 1036 AAAGGAAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1095
QY 961 CTGCGCCCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db 1096 CTGCGCCCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1155
QY 1021 CGGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1080
Db 1156 CGGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1215
QY 1081 CTGCGGACGAGTTCTGATT 1100
Db 1216 CTGCGGACGAGTTCTGATT 1235
```

## RESULT 4

```
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/774,954
FILING DATE: 30-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-Nov-1997
APPLICATION NUMBER: 08/792,498
FILING DATE: 31-Jan-1997
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 11284 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5
```

```
Query Match 99.9%; Score 1098.4; DB 11; Length 11284;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCCCGGGCTCTGGGACCCGCGGCTTACTGCTCTACTGCCCCCTGCATGGGGCGC 60
Db 4236 ATGCCCGGGCTCTGGGACCCGCGGCTTACTGCTCTACTGCCCCCTGCATGGGGCGC 4295
QY 61 TTGGGAACACGCGCGATCACTCTTGGCTCTTGGCATTTGCAAGCTGTAACCGT 120
Db 4296 TTGGGAACACGCGCGATCACTCTTGGCTCTTGGCATTTGCAAGCTGTAACCGT 4355
QY 121 ACCTTGGTGTCTCTCTTGGATTTGAGTACAGGATCAAGGCTCTCTTTCACCACTC 180
Db 4356 ACCTTGGTGTCTCTCTTGGATTTGAGTACAGGATCAAGGCTCTCTTTCACCACTC 4415
QY 181 CATGTGCTCTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
Db 4416 CATGTGCTCTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 4475
QY 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACCTGGCCCTCGAGAAGCGGGTG 300
Db 4476 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACCTGGCCCTCGAGAAGCGGGTG 4535
QY 301 GCATACCTGCTTTGAGTGGACCCAGCGAAGCCAGATGAAGACGTCGCCCATGAAG 360
Db 4536 GCATACCTGCTTTGAGTGGACCCAGCGAAGCCAGATGAAGACGTCGCCCATGAAG 4595
QY 361 GAAGGAACACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db 4596 GAAGGAACACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 4655
QY 421 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAACCAATGGAGCCAGAGATTTTCT 480
Db 4656 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAACCAATGGAGCCAGAGATTTTCT 4715
QY 481 CCAAGGAACATCCGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 4716 CCAAGGAACATCCGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4775
QY 541 GAACACAGGCACTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 4776 GAACACAGGCACTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4835
```



QY 1021 CGGAGCGGACCTCAGGGAGGCGCTCTCTTTCTTGGCATGAGACAGGCCCTTAAG 1080  
Db 1139 CGGAGCGGACCTCAGGGAGGCGCTCTCTTTCTTGGCATGAGACAGGCCCTTAAG 1198  
QY 1081 CTGGGACAGTTCGTGATT 1100  
Db 1199 CTGGGACAGTTCGTGATT 1218

RESULT 6  
; Sequence 4752, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956,157  
; NUMBER OF SEQ ID NOS: 2004-10-04  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4752  
; LENGTH: 5218  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-4752

Query Match 99.7%; Score 1096.8; DB 22; Length 5218;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1098; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCCCGGGCTCTGGGACCGCGCGTGTACTGCTCTACTGCGCCCTGATGGGCGC 60  
Db 119 ATGCCCGGGCTCTGGGACCGCGCGTGTACTGCTCTACTGCGCCCTGATGGGCGC 178  
QY 61 TTGGAACAGCGGACCTCACTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAAGCT 120  
Db 179 TTGGAACAGCGGACCTCACTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAAGCT 238  
QY 121 ACCTTGGTGTCTCTCTTGGATGAGTACAGCATCACAAGCTCTCTTCAACAACCTC 180  
Db 239 ACCTTGGTGTCTCTCTTGGATGAGTACAGCATCACAAGCTCTCTTCAACAACCTC 298  
QY 181 CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCGCCCTCCAGGCTTACCATCGGGTCATC 240  
Db 299 CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCGCCCTCCAGGCTTACCATCGGGTCATC 358  
QY 241 AGCTTGGAGGATTCATGGAGAGCTGGCACCCACCTGCGCCCTGAGAGCGGGTG 300  
Db 359 AGCTTGGAGGATTCATGGAGAGCTGGCACCCACCTGCGCCCTGAGAGCGGGTG 418  
QY 301 GCATAGCTCTTGGTGGAGCGCCAGCGAGCCAGATGAAGACGTGCCCTCATGAAG 360  
Db 419 GCATAGCTCTTGGTGGAGCGCCAGCGAGCCAGATGAAGACGTGCCCTCATGAAG 478  
QY 361 GAAGGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 479 GAAGGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538  
QY 421 CTTTTTCAGGCAATTTCTTCAAGTCTTCTTACAGAACTTGGAGCGAGAGATTTTCT 480  
Db 539 CTTTTTCAGGCAATTTCTTCAAGTCTTCTTACAGAACTTGGAGCGAGAGATTTTCT 598  
QY 481 CCNAGGAACTCGGTGCTTGGCTGCGAGGCGCCAGCCAGTCCCGTCTAGAA 540  
Db 599 CCNAGGAACTCGGTGCTTGGCTGCGAGGCGCCAGCCAGTCCCGTCTAGAG 658  
QY 541 GAACACAGGCCACTACAGAGATCATGTGTATGGTTCAGCGAAATGGTGAAGACGGGAGAG 600  
Db 659 GAACACAGGCCACTACAGAGATCATGTGTATGGTTCAGCGAAATGGTGAAGACGGGAGAG 718

## RESULT 7

US-09-774-954-7  
; Sequence 7, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498  
; FILING DATE: 31-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041PI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881





```

; Publication No. US20050118625A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Wyeth
;
; APPLICANT: Mounts, William
;
; TITLE OF INVENTION: NUCLEOTIC ACID SEQUENCES
;
; TITLE OF INVENTION: HUMAN OST
;
; FILE REFERENCE: 031896-043000
;
; CURRENT APPLICATION NUMBER: US
;
; CURRENT FILING DATE: 2004-10-
;
; NUMBER OF SEQ ID NOS: 319805
;
; SOFTWARE: PatentIn version 3.2.
;
; SEQ ID NO 4751
;
; LENGTH: 1506
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; US-10-956-157-4751

```

| Query Match           | 42.9%;          | Score 471.4;   | DB 22;    | Length 1506; |
|-----------------------|-----------------|--|-----------|--------------|
| Best Local Similarity | 99.8%;          | Prod. No. 1.9e-137;  |           |              |
| Matches 472;          | Conservative 0; | Mismatches 1;  | Indels 0; | Gaps 0;      |
| Qy                    | 1               | ATGCCCGCGGCTCTCTGGAGCCCGCGGTTACTCTCTACTCGCCCTGCATGGGGGCG       | 60        |              |
| Db                    | 119             | ATGCTCGCGGCTCTCTGGAGCCCGCGGTTACTCTCTACTCGCCCTGCATGGGGGCG       | 178       |              |
| Qy                    | 61              | TTTGGGAACACAGCCGATCACTCTTTCGGGCTCTCTGGCAATTTGCAAGCTGCTAAACGCT  | 120       |              |
| Db                    | 179             | TTTGGGAACACAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACGCT   | 238       |              |
| Qy                    | 121             | ACCTTGGCTGCTCCCTTGGATTGAGTACACAGCATCAACAGCTCTCTTTCACCAACCTC    | 180       |              |
| Db                    | 239             | ACCTTGGCTGCTCCCTTGGATTGAGTACACAGCATCAACAGCTCTCTTTCACCAACCTC    | 298       |              |
| Qy                    | 181             | CATGTGTCCTACACAGAAGTACTTCAAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC  | 240       |              |
| Db                    | 299             | CATGTGTCCTACACAGAAGTACTTCAAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC  | 358       |              |
| Qy                    | 241             | AGCTTTGAGGATTTTCATGAGAGAAGCTGGCACCCACCACTGSCCCCTCGAAGAAGCGGGTG | 300       |              |
| Db                    | 359             | AGCTTTGAGGATTTTCATGAGAGAAGCTGGCACCCACCACTGSCCCCTCGAAGAAGCGGGTG | 418       |              |
| Qy                    | 301             | GCATACTGCTTTGAGGTGTCAGCCAGCGAGCCAGATAGAGAGAGTGTCGCCCATGAAG     | 360       |              |
| Db                    | 419             | GCATACTGCTTTGAGGTGTCAGCCAGCGAGCCAGATAGAGAGAGTGTCGCCCATGAAG     | 478       |              |
| Qy                    | 361             | GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG | 420       |              |
| Db                    | 479             | GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG | 538       |              |
| Qy                    | 421             | CTTTTTCACGCCATTTCCCTTCAGTGCTTCTCCTACAGAGAACAATGGAGCCAGAG       | 473       |              |
| Db                    | 539             | CTTTTTCACGCCATTTCCCTTCAGTGCTTCTCCTACAGAGAACAATGGAGCCAGAG       | 591       |              |

```

RESULT 10
US-09-764-891-6032
; Sequence 6032, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Pr
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - cons
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6032
; LENGTH: 10331
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6032

```

|                       |        |              |             |            |     |        |        |
|-----------------------|--------|--------------|-------------|------------|-----|--------|--------|
| Query Match           | 22.8%; | Score        | 251.2;      | DB         | 10; | Length | 10331; |
| Best Local Similarity | 97.0%; | Pred.        | No.9.4e-66; |            |     |        |        |
| Matches               | 256;   | Conservative | 0;          | Mismatches | 8;  | Indels | 0;     |
| Gaps                  | 0;     |              |             |            |     |        |        |

  

|    |      |          |           |         |          |          |          |          |           |            |      |
|----|------|----------|-----------|---------|----------|----------|----------|----------|-----------|------------|------|
| QY | 656  | GCTCTGAC | TGGGAAGAC | CCCTGTG | CCCATCTG | CTAAGGAC | GGGACTG  | CAGCGCT  | CGCACTTCA | 715        |      |
|    |      |          |           |         |          |          |          |          |           |            |      |
| Db | 2480 | GCTCTCCG | TAGAAGAC  | CCCTGTG | CCCATCTG | CTAAGGAC | GGGACTG  | CAGCGCT  | CGCACTTCA | 2539       |      |
|    |      |          |           |         |          |          |          |          |           |            |      |
| QY | 716  | TGGCCTCT | CCGCA     | GTGTGTG | GGGTAC   | AGCCG    | CAGCAC   | AGCGGCC  | CCCCCTC   | ACGATGACTA | 775  |
|    |      |          |           |         |          |          |          |          |           |            |      |
| Db | 2540 | TGGCCTCT | CCGCA     | GTGTGTG | GGGTAC   | AGCCG    | CAGCAC   | AGCGGCC  | CCCCCTC   | ACGATGACTA | 2599 |
|    |      |          |           |         |          |          |          |          |           |            |      |
| QY | 776  | TGTGCTCT | GCCTC     | CACTGA  | AGAGATC  | CAGAGGG  | CTGTG    | AAAGCTCT | CGGTGAGGT | CGCTGG     | 835  |
|    |      |          |           |         |          |          |          |          |           |            |      |
| Db | 2600 | TGTGCTCT | GCCTC     | CACTGA  | AGAGATC  | CAGAGGG  | CTGTG    | AAAGCTCT | CGGTGAGGT | CGCTGG     | 2659 |
|    |      |          |           |         |          |          |          |          |           |            |      |
| QY | 836  | ATGCCAGT | TCGGTCT   | ACGTTG  | CTACTG   | ATTTCG   | AGAGTTAT | TGCTG    | AGCTTCC   | AAACAGC    | 895  |
|    |      |          |           |         |          |          |          |          |           |            |      |
| Db | 2660 | ATGCCAGT | TCGGTCT   | ACGTTG  | CTACTG   | ATTTCG   | AGAGTTAT | TGCTG    | AGCTTCC   | AAACAGC    | 2719 |
|    |      |          |           |         |          |          |          |          |           |            |      |
| QY | 896  | TCTTCAA  | AGGGAAG   | TGAAG   | TGGTGG   | 919      |          |          |           |            |      |
|    |      |          |           |         |          |          |          |          |           |            |      |
| Db | 2720 | TCTTCAA  | AGGGAAG   | TGAAG   | TGGTGG   | 2743     |          |          |           |            |      |
|    |      |          |           |         |          |          |          |          |           |            |      |

  

```

RESULT 11
US-09-918-995-19437
Sequence 19437, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19437
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-19437

```

| Query Match           | 21.2%          | Score 233  | DB 10    | Length 479 |
|-----------------------|----------------|--|----------|------------|
| Best Local Similarity | 100.0%         | Prod. No. 1.7e-62  |          |            |
| Matches 233           | Conservative 0 | Mismatches 0   | Indels 0 | Gaps 0     |
| Qy                    | 868            | GAGAGTTATGTGCTCGACTCCAAACAGCTCTTCAAAGGGAAGGTGAGTGTGAGCCTG    | 927      |            |
| Db                    | 31             | GAGAGTTATGTGCTCGACTCCAAACAGCTCTTCAAAGGGAAGGTGAGTGTGAGCCTG    | 90       |            |
| Qy                    | 928            | AAGCTGAGGTGGCCAGGTCGACCTGTATCATCTTCGCCCAAGCCGACCACTTTATTGGC  | 987      |            |
| Db                    | 91             | AAGCCTGAGGTGGCCAGGTCGACCTGTATCATCTTCGCCCAAGCCGACCACTTTATTGGC | 150      |            |
| Qy                    | 988            | AACGTGTCTCTCTCTTCATCTGCCTTTGTGAAGCGGGACGGGACCTTCAGGGGAGGCCG  | 1047     |            |
| Db                    | 151            | AACGTGTCTCTCTCTTCATCTGCCTTTGTGAAGCGGGACGGGACCTTCAGGGGAGGCCG  | 210      |            |
| Qy                    | 1048           | TCCTCTTTCTTCGGCATCGGACAGGCCCCCTTAAGCTGCGGGACAGTTCTGATT       | 1100     |            |
| Db                    | 211            | TCCTCTTTCTTCGGCATCGGACAGGCCCCCTTAAGCTGCGGGACAGTTCTGATT       | 263      |            |

RESULT 12  
US-09-918-995-22744  
; Sequence 22744, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.

```
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: FROM VARIOUS CDNA LIBRARIES
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22744
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(477)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22744
```

```
Query Match      20.2%; Score 221.8; DB 10; Length 477;
Best Local Similarity 96.3%; Pred. No. 5.9e-59;
Matches 237; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY      855 TCCTACTGATCCCGAGCTTATGTGCGCTGAGCTCCACAGCTCTTCAAAGGGAAGGTGAA 914
Db      11  TGATCTGAGCNCTAGAGATATGTGCTGAGCTCC-ACAGTTCTTCAAAGGGAAGGTGAA 69

QY      915 GGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATCTCTCGGCGAAGCCGA 974
Db      70  GGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATCTCTCGGCGAAGCCGA 129

QY      975 CCACTTTATGGCAACTGTCTCTCTCTCACTGCTCTTGTGAAGCGGAGCGGACCT 1034
Db      130 CCACTTTATGGCAACTGTCTCTCTCTCTCACTGCTCTTGTGAAGCGGAGCGGACCT 189

QY      1035 CCAGGGAGGCGCTTCTTCTTCTGCGCATGAGCAGGCCCCCTAAGCTGCGGAGCAGTT 1094
Db      190 CCAGGGAGGCGCTTCTTCTTCTGCGCATGAGCAGGCCCCCTAAGCTGCGGAGCAGTT 249

QY      1095 CTGATT 1100
Db      250 CTGATT 255
```

```
RESULT 13
US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
```

```
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18161
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18161
```

```
Query Match      15.1%; Score 166.4; DB 26; Length 1209;
Best Local Similarity 51.5%; Pred. No. 2.4e-41;
Matches 563; Conservative 0; Mismatches 486; Indels 45; Gaps 6;

QY      17  GGGACCCGCGCGTTTACCTGCTCTACTGCCCTGTCATGGGGCGCTTTGGGAACCAAGCGCG 76
Db      80  GCGATCCCAATGGCTACCTACTCTCGTGTATGGAGCGCTTTGGCAACCAAGCGCG 139

QY      77  ATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGTCTAAACGTTACCTTTGGCTGTCCCTC 136
Db      140 ACCACTTCTTGGGATCATTTGGCCTTCGCCAAGGCGCTTTAATCGCACCTGTATCTCCGCG 199

QY      137 CTTGATTTGATACCAGCATCACAGCCTCTTTTACCACCACTCCATGTCCTTACCAGA 196
Db      200 CGTGGTGGAGT-----ATCGTAGGGGTGAATCGCGATCCCGTCAGGTACCGTTCAACA 253

QY      197 AGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTTCATCAGCTTTGGAGGATTTCA 256
Db      254 CATATTTTGAAGTGGAGCCCTCGAAGGAATACCATCGGTCATCACCATGGCAGATTTCA 313

QY      257 TGGAGAGCTGGACCCACCCACTGGCCCCCTGAGAACGGGGTGGCATACTGCTTTTGAGG 316
Db      314 TGTGGCACCTGGCGGACGACATTTTGGCCAGAATCGGAGCGAGTGTCTATTTTGTCTACAAAG 373

QY      317 -----TGGCAGCCCGAGGAGCCGAGATAAGAGACGTCGCCCATGAAGG 361
Db      374 AACGATATAGCCTTCAGCAGGAGAGAGAACGATTCAGACAAAGCCCAATTGCACGCGCAAG 433

QY      362 AAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAGTCGGAGC 421
Db      434 ATGGCAATCCTTTTGTGTCCTTTTGGGACACTTTTTCATATAGACTTTTGTGCGGTGAGGT 493

QY      422 TTTTACAGGCATTTCTTTCAGTGTCTCTTACAG-----AGAACATGGAGCCAGA 472
Db      494 TCTATGGCCCACTTCAATTTTGTATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCA 553

QY      473 GATTTTCTCAAAGGAACATCCGCTGCTTGGCTTGGCCCTGCCAGAGGAGCCCGCAGCTTCCCG 532
Db      554 AATATCCTGCGAGAAATCATATCCCGTACTCGCGTTTTCACCGGAGCTCCGCTAGTTTTCCTG 613

QY      533 TCCTAGAAGAACACAGGCGCACTACAGAAAGTACATGGTATGTCAGACGAAATGGTGAAGA 592
Db      614 TTCAGCTAGAGAACTGCAAGCTGCGAGCGCTACTTTCAGTGGAGTCAACGGTATAGGGAAG 673

QY      593 CGGAGAGGCCAGATTCATGCCCCACCTTGTGTCGCGC---CCTATGTGGGCAATTCATCTGC 649
Db      674 CATCTAAGGATTTTCATCCGAGAGCAGTTTGCCTCGGGGTGCGCTTTTGGGCAATTCATCTGC 733

QY      650 GCATTTGCTCTGACTGGAAGAACCCCTGTGCCATGCTCTGAAGGACGG-----GACTG 700
Db      734 GCAACGGTATCGATTGGGTGAGAGCCCTGTGAGCAGCTCAAGGATAGCCAGCATCTGTGTTG 793

QY      701 CAGGCTCGCACTTCATGGCCCTCTCCGCGCTGTGTGGGCTACAGCGCGAGCAGCGGCGCC 760
Db      794 CCTCGCGCAGTGTCTGGGCTATAAAATGAACTGGTGTGCACTTACCCGAGGCTCTGCA 853

QY      761 CCCTCAGATGATATGTGCTGCTGCTGACCTTGAAGGAGATCCAGAGGCTGTGAAGCTCT 820
Db      854 TGCCCTCAAAGGAGGCGCATCATCCGCCAGCTTAAAGAGAACCATTAAGAACCTGCGCCAAA 913

QY      821 GGGTGAGTCTGCTGGATGCCCGCAGTCGCTCTACGTTGCTACTGATTCGAGAGTATGTGC 880
Db      914 CTCAGCCGGACAAAGAAATCAAAATCAGTTTTCGTGGCGTCCAGACTTCAATCACATGATTG 973

QY      881 CTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGGCTGGAAGCTGAGGTGG 940
```

Db 974 GTGAACAAACAGGCGCTTAGTCGATGGGCATAGTGTGCAACAGCTGCGGAGGATG 1033  
Qy 941 ---CCAGGTGCGACCTGTATCCTCGCCCAAGCCGACCACTTTATTTGGCAACTGTGTCT 997  
Db 1034 ATCCTTACCTGGACTTTGGCCATCTCGACAGTGGAAACCACTTTATTCGGCAACTGTATAT 1093  
Qy 998 CTTCTCTTCACTGCTTTGTGAAGCGGAGCGGACCTTCCAGGGGAGGCGCTTCTTCTTCT 1057  
Db 1094 CTTCTTACTCGGCATTCGAAAAAGGAAACGAGATGTGCACGGTTTTTCATCGTACTTCT 1153  
Qy 1058 TCGGCATGGACAGG 1071  
Db 1154 GGGGATCCCCCAG 1167

RESULT 14

US-11-097-143-18160/c  
; Sequence 18160, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,931  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18160  
; LENGTH: 3264  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-18160

Query Match 13.7%; Score 150.4; DB 26; Length 3264;  
Best Local Similarity 50.9%; Pred. No. 3.7e-36;  
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;

Qy 49 TGCATGGGCGCTTTGGGAACAGGCGCATCTCTTGGGCTCTCTGGCATTTGCAAG 108  
Db 2098 TTCAAGAGCGCTTTGGCAACAGGCGGACCATCTCTGGGATCATTTGGCTTCGCCAAG 2039  
Qy 109 CTGCTAACCCTGACTTGGCTGTCCTCTGGAATTTGAGTACAGCATCACAAGCTTCT 168  
Db 2038 GCGCTTAATCGCACCTTGATCTTCGCGCGCTGGGGGAGT-----ATCtRAGGGTGA 1985  
Qy 169 TTCACCAACCTTCATGTGTCTTACACGAAGTACTTTCAAGCTGGAGCCCTCCAGGCTTAC 228  
Db 1984 CTGGATCCCGTCAAGTACCGTGTCAACACATATTTGAAGTGGAGCCCTCGAAGGAATAC 1925  
Qy 229 CATCGGCTCATCAGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCCACCTGGCCCCCT 288  
Db 1924 CATCGCTCATCACCATGGCAGATTTTCATGTGGCACCTGGCCGACGACATTTTGGCCAGAA 1865

Qy 289 GAGAAAGCGGTGGGCATATCTGCTTTGAGG-----TGAGCCGCCAGCAAGC 333  
Db 1864 TCGGAGCGAGTGTATCTTTGCTACAAGGAACGATATAGCCTTCAGCAGGAGAAGCAAT 1805  
Qy 334 CCAGATAGAGAGAGCTGCCCCCATTAAGAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAG 393  
Db 1804 CCAGCAAGCCCAATTTGCCACGCCAAGATGCAATCTTTTGGTCCCTTTTGGGACACT 1745  
Qy 394 TTTTCAATGTAGTGTTCACAAGTCCGAGCTTTTTCAGAGCATTTTCTTTCAGTGTCTTCTTAC 453  
Db 1744 TTTTCAATAGACTTTTGTGGGTTCAGAGTTCTATGCCCACTTCATTTTGTATGTGCATCAT 1685  
Qy 454 AG-----AGAACAAATGGAGCCAGAGATTTTCTCCAAAGGAACATCCGGTGTCTGCC 504  
Db 1684 AGCAACGAGGCTGCCAAGTGGCAGACCAATATCTTCGAGATCATATCCCTACTCGG 1625  
Qy 505 CTGCCAGGAGCCCGCAGCCAGTTCCTCCCTCTAGAGAAACACAGGCGCATACAGAAGTAC 564  
Db 1624 TTTACCGGAGCTCCGGCTAGTTTTCTGTTAGTGAAGAACTGCAAGCTGCGGCTTAC 1585  
Qy 565 ATGATATGTCAGACGAATATGTTGAAGACGGGAGAGCCCGAGATTTATGCCACCTTGTTC 624  
Db 1564 TTGCAGTGGAGTCAACGGTATAGGGAAGCATCTTAAGGATTTTCATCCGAGAGCAGTTGCT 1505  
Qy 625 CGG---CCCTATGTGGGCATTTCTGCGCATTTGGCTCTGACTGGAAGAACCGCTGTGCC 681  
Db 1504 CGGGGTGCTTTTGGGCAATTCATCTGCGCAACGATATCGATTTGGGTGAGAGCCTGTGAG 1445  
Qy 682 ATGCTGAAGAGACGG-----GACTGAGGCTCGCACTTTCATGGCCCTCTCCGAGTGT 732  
Db 1444 CACGTCAAGGATAGCCAGCATCTGTTTGCCTCGCGCAGTCTTGGGCTATATAAATGAA 1385  
Qy 733 GTGGGCTACAGCGCAGCAGCAGCGGCCCTCTCAGATGACTATGTGCTGCTGCTGACCTG 792  
Db 1384 CGTGTGCACCTTACCCGGAGCTCTGCATGCCCTCAAGGAGGCGATCATCCCGCAGCTA 1325  
Qy 793 AAGGAGATCCAGAGGCGCTGGAAGCTCTGGGTGAGTCTGGTGTGAGTCCAGTCCAGTCTGCTAC 852  
Db 1324 AAGAGAACCATTAAGAACCTGCGCCAACTCAGCCGCAACGAATCAATCAGTTTC 1265  
Qy 853 GTTCTACTGATTCGAGAGTATATGCTGAGTCTCAACAGCTCTTCAAGGGAGGTG 912  
Db 1264 GTGGGTGAGTCTCCATCATGATTTGGTGAATTAACACGCGCTTGTAGTCGATGGC 1205  
Qy 913 AAGTGTGAGCTGGAAGCTGAGGTGG---CCAGGTGCACTGTACATCTCTCGGCCA 969  
Db 1204 ATCAGTGTGCAAGCTGCGGAGGATGATCTTACCTGGACTTGGCATTTCTCGGACAG 1145  
Qy 970 GCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCTTCTTGTGAAGCGGAGCGG 1029  
Db 1144 TCGAACCACTTTATCGGCAACTGTATATCTTCTTACTCGGCATTCGAAAAAAGGGAACGA 1085  
Qy 1030 GACCTCCAGGCGAGCGCTCTTCTTTCTTGGCATGGACAGG 1071  
Db 1084 GATGTGCACGGTTTTTCCATCGTACTTCTGGGATTTCCCAAG 1043

RESULT 15

US-09-783-590-11501  
; Sequence 11501, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12

```

; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11501
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (333)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (342)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (400)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-11501

```

```

Query Match      10.4%; Score 114.6; DB 9; Length 402;
Best Local Similarity 83.9%; Pred. No. 3.4e-25;
Matches 177; Conservative 0; Mismatches 23; Indels 11; Gaps 4;

Qy 178 CTCATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTC 237
Db 161 CTCATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTC 220

Qy 238 ATCAGCTT-GGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGCCCCCT---GAGAA 293
Db 221 ATCAGCTTGGGGAGATTTTCATGGAGAGCTGGCACCACCCACTGGCCCCCTGAGAAGC 280

Qy 294 GCGGGTGGCATACTGCTTTGA--GGTGGCAGCCAGCAAGCCAGAT-----AAGAAGA 346
Db 281 GGGGTGGGCATATGCTTTTNGTGGGCGCCAGCCGCGGCGCCAGATTAAGNAGGACG 340

Qy 347 CGTGCCCATGAAGAGCAACCCCTTTGG 377
Db 341 TNGACCCCATGAAGAGCAACCCCTTTG 371

```

Search completed: November 9, 2005, 15:51:17  
Job time : 988.698 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 02:15:54 ; Search time 199.044 Seconds  
(without alignments)  
9042.764 Million cell updates/sec

Title: US-09-774-954-16  
Perfect score: 1100  
Sequence: 1 ATGCCCGCGGCTCTCTGGGA.....CTCGCGGACGAGTCTTGATT 1100

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/5A-COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/5B-COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PTUS-COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length  | ID | Description         |
|------------|--------|-------------|---------|----|---------------------|
| 1          | 1100   | 100.0       | 1100    | 3  | US-08-978-741-16    |
| 2          | 1100   | 100.0       | 1100    | 3  | US-09-333-729A-16   |
| 3          | 1100   | 100.0       | 1514    | 3  | US-08-978-741-1     |
| 4          | 1100   | 100.0       | 1514    | 3  | US-09-333-729A-2    |
| 5          | 1098.4 | 99.9        | 1300    | 3  | US-08-978-741-4     |
| 6          | 1098.4 | 99.9        | 1300    | 3  | US-09-333-729A-6    |
| 7          | 1098.4 | 99.9        | 11284   | 3  | US-08-978-741-5     |
| 8          | 1033.4 | 93.9        | 5009    | 3  | US-08-978-741-7     |
| 9          | 1033.4 | 93.9        | 5009    | 3  | US-09-333-729A-8    |
| 10         | 168    | 15.3        | 1320    | 4  | US-09-270-767-14353 |
| 11         | 62.2   | 5.7         | 722     | 4  | US-09-270-767-30507 |
| 12         | 58.6   | 5.3         | 998     | 4  | US-09-270-767-28905 |
| 13         | 58.6   | 5.3         | 2034    | 4  | US-09-270-767-13022 |
| 14         | 56     | 5.1         | 19227   | 4  | US-09-949-016-12127 |
| 15         | 56     | 5.1         | 19228   | 4  | US-09-949-016-12127 |
| 16         | 41.2   | 3.7         | 114793  | 4  | US-10-148-806-3     |
| 17         | 40     | 3.6         | 40      | 3  | US-08-978-741-10    |
| 18         | 40     | 3.6         | 40      | 3  | US-08-978-741-12    |
| 19         | 40     | 3.6         | 40      | 3  | US-09-333-729A-9    |
| 20         | 39     | 3.5         | 1131    | 4  | US-09-252-991A-76   |
| 21         | 39     | 3.5         | 1353    | 4  | US-09-252-991A-88   |
| 22         | 38.6   | 3.5         | 2358    | 4  | US-09-919-039-190   |
| 23         | 37.8   | 3.4         | 2154    | 4  | US-09-902-540-4557  |
| 24         | 37.8   | 3.4         | 24905   | 4  | US-09-902-540-4557  |
| 25         | 37.8   | 3.4         | 4403765 | 3  | US-09-103-840A-2    |
| 26         | 37.8   | 3.4         | 4411529 | 3  | US-09-103-840A-1    |
| 27         | 37.4   | 3.4         | 18037   | 4  | US-09-949-016-13678 |

|      |      |     |        |   |                      |                    |
|------|------|-----|--------|---|----------------------|--------------------|
| C 28 | 37.4 | 3.4 | 77536  | 4 | US-09-410-551B-1     | Sequence 1, Appli  |
| C 29 | 37.4 | 3.4 | 77536  | 4 | US-09-940-316B-1     | Sequence 1, Appli  |
| C 30 | 37.2 | 3.4 | 933    | 4 | US-09-252-991A-11943 | Sequence 11943, A  |
| C 31 | 37.2 | 3.4 | 1218   | 4 | US-09-252-991A-12011 | Sequence 12011, A  |
| C 32 | 37.2 | 3.4 | 1650   | 4 | US-09-252-991A-11977 | Sequence 11977, A  |
| C 33 | 37.2 | 3.4 | 102053 | 4 | US-09-949-016-13025  | Sequence 13025, A  |
| C 34 | 37   | 3.4 | 1843   | 1 | US-07-918-023-1      | Sequence 1, Appli  |
| C 35 | 37   | 3.4 | 5962   | 6 | 5386025-5            | Patent No. 5386025 |
| C 36 | 37   | 3.4 | 5962   | 6 | 5386025-5            | Patent No. 5386025 |
| C 37 | 37   | 3.4 | 5975   | 1 | US-08-404-354B-1     | Sequence 1, Appli  |
| C 38 | 37   | 3.4 | 5975   | 1 | US-08-314-083B-1     | Sequence 1, Appli  |
| C 39 | 37   | 3.4 | 5975   | 1 | US-08-435-675B-1     | Sequence 1, Appli  |
| C 40 | 37   | 3.4 | 5975   | 1 | US-08-336-257A-3     | Sequence 3, Appli  |
| C 41 | 37   | 3.4 | 5975   | 3 | US-08-884-599-1      | Sequence 1, Appli  |
| C 42 | 36.8 | 3.3 | 3591   | 4 | US-10-237-551-191    | Sequence 191, App  |
| C 43 | 36.8 | 3.3 | 3591   | 4 | US-10-237-551-221    | Sequence 221, App  |
| C 44 | 36.8 | 3.3 | 3981   | 4 | US-09-294-298A-5     | Sequence 5, Appli  |
| C 45 | 36.8 | 3.3 | 3981   | 4 | US-09-294-298A-24    | Sequence 24, Appli |

ALIGNMENTS

RESULT 1  
US-08-978-741-16  
; Sequence 16, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1100 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear

US-08-978-741-16  
Query Match 100.0%; Score 1100; DB 3; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCCCGCGGCTCTCTGGACCCCGGTTACTGCTCTACTGCCCTGATGGGCGC 60  
DB 1 ATGCCCGCGGCTCTCTGGACCCCGGTTACTGCTCTACTGCCCTGATGGGCGC 60

Qy 61 TTTGGGAACAGGCGCATCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACAGGCGCATCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGCATCACAAGCTCTCTTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGCATCACAAGCTCTCTTTACCAACCTC 180  
Qy 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240  
Db 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240  
Qy 241 AGCTTGGAGTTTCATGAGAGCTGCACCCACCTCTGGCCCTCTGAGAGCGGGTG 300  
Db 241 AGCTTGGAGTTTCATGAGAGCTGCACCCACCTCTGGCCCTCTGAGAGCGGGTG 300  
Qy 301 GCATCTGCTTGGAGTGGAGCCAGCGAAGCCAGATGAAGAAGCGTGCCTCATGAAG 360  
Db 301 GCATCTGCTTGGAGTGGAGCCAGCGAAGCCAGATGAAGAAGCGTGCCTCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTATTACAGGCAATTCCTTCACTGCTTCTTACAGAGAAACAATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTATTACAGGCAATTCCTTCACTGCTTCTTACAGAGAAACAATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGGTGTCTGCCCTGCCAGGAGCCCGCCAGTCCCGCTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTCTGCCCTGCCAGGAGCCCGCCAGTCCCGCTCTAGAA 540  
Qy 541 GAAACACAGGCCACTACAGAGTACATGTATGTATGTCAGAGAAATGTAAGACGGGAG 600  
Db 541 GAAACACAGGCCACTACAGAGTACATGTATGTATGTCAGAGAAATGTAAGACGGGAG 600  
Qy 601 GCCAGATTCATGCCCACTCTGTCGGCCCTATGTGGGATTCATCTGGGCAATTTGGCTCT 660  
Db 601 GCCAGATTCATGCCCACTCTGTCGGCCCTATGTGGGCAATTCATCTGGGCAATTTGGCTCT 660  
Qy 661 GACTGGAAGAACCGCTGTGCAATGTGAAGACGGGACTGAGGCTCGCTGATGTC 720  
Db 661 GACTGGAAGAACCGCTGTGCAATGTGAAGACGGGACTGAGGCTCGCTGATGTC 720  
Qy 721 TCTCCGAGTGTGTGGGCTACAGCGGCTGTGAAGCTCTGGGTGAGGTGCGCTGGATGCC 840  
Db 721 TCTCCGAGTGTGTGGGCTACAGCGGCTGTGAAGCTCTGGGTGAGGTGCGCTGGATGCC 840  
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGCTGGATGCC 840  
Db 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGCTGGATGCC 840  
Qy 841 CAGTGGTCTACGTTGCTACTGATTCAGAGAGTTATGCTGAGCTCCACAGCTCTTC 900  
Db 841 CAGTGGTCTACGTTGCTACTGATTCAGAGAGTTATGCTGAGCTCCACAGCTCTTC 900  
Qy 901 AAAGGGAAGTGAAGTGTGAGCTGAAGCTGAGGTGCGCCAGGTGGAAGCTGTATC 960  
Db 901 AAAGGGAAGTGAAGTGTGAGCTGAAGCTGAGGTGCGCCAGGTGGAAGCTGTATC 960  
Qy 961 CTCGGCCAGCCGACACATTTATTTGGCAACTGTGTCTCTCTTCACTGCTCTTTGAG 1020  
Db 961 CTCGGCCAGCCGACACATTTATTTGGCAACTGTGTCTCTCTTCACTGCTCTTTGAG 1020  
Qy 1021 CGGAGGCGGACCTCCAGGGAGGCGCTCTTCTTTCTTGGCATGACAGGCGCCCTAAG 1080  
Db 1021 CGGAGGCGGACCTCCAGGGAGGCGCTCTTCTTTCTTGGCATGACAGGCGCCCTAAG 1080  
Qy 1081 CTGCGGACGAGTTCTGATT 1100  
Db 1081 CTGCGGACGAGTTCTGATT 1100

RESULT 2  
US-09-333-729A-16  
; Sequence 16, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Rucosyltransferase  
; FILE REFERENCE: P1041PDI1-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 16  
; LENGTH: 1100  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-333-729A-16  
  
Query Match 100.0%; Score 1100; DB 3; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ATGCCCGGGGCTCTCTGGACCCGCGGTTACCTGCTCTACTGCCCTGTCATGGGGCG 60  
Db 1 ATGCCCGGGGCTCTCTGGACCCGCGGTTACCTGCTCTACTGCCCTGTCATGGGGCG 60  
Qy 61 TTTGGGAACAGGCGCATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACAGGCGCATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGCATCACAAGCTCTCTTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGCATCACAAGCTCTCTTTACCAACCTC 180  
Qy 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240  
Db 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240  
Qy 241 AGCTTGGAGTTTCATGAGAGCTGCACCCACCTCTGGCCCTCTGAGAGCGGGTG 300  
Db 241 AGCTTGGAGTTTCATGAGAGCTGCACCCACCTCTGGCCCTCTGAGAGCGGGTG 300  
Qy 301 GCATCTGCTTGGAGTGGAGCCAGCGAAGCCAGATGAAGAAGCGTGCCTCATGAAG 360  
Db 301 GCATCTGCTTGGAGTGGAGCCAGCGAAGCCAGATGAAGAAGCGTGCCTCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTATTACAGGCAATTCCTTCACTGCTTCTTACAGAGAAACAATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTATTACAGGCAATTCCTTCACTGCTTCTTACAGAGAAACAATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGGTGTCTGCCCTGCCAGGAGCCCGCCAGTCCCGCTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTCTGCCCTGCCAGGAGCCCGCCAGTCCCGCTCTAGAA 540  
Qy 541 GAAACACAGGCCACTACAGAGTACATGTATGTATGTCAGAGAAATGTAAGACGGGAG 600  
Db 541 GAAACACAGGCCACTACAGAGTACATGTATGTATGTCAGAGAAATGTAAGACGGGAG 600  
Qy 601 GCCAGATTCATGCCCACTCTGTCGGCCCTATGTGGGCAATTCATCTGGGCAATTTGGCTCT 660  
Db 601 GCCAGATTCATGCCCACTCTGTCGGCCCTATGTGGGCAATTCATCTGGGCAATTTGGCTCT 660  
Qy 661 GACTGGAAGAACCGCTGTGCAATGTGAAGACGGGACTGAGGCTCGCTGATGTC 720  
Db 661 GACTGGAAGAACCGCTGTGCAATGTGAAGACGGGACTGAGGCTCGCTGATGTC 720  
Qy 721 TCTCCGAGTGTGTGGGCTACAGCGGCTGTGAAGCTCTGGGTGAGGTGCGCTGGATGTC 780





QY 1081 CTGCGGACGAGTTCTGATT 1100  
Db 1081 CTGCGGACGAGTTCTGATT 1100

## RESULT 4

US-09-333-729A-2  
; Sequence 2, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P1041PDI-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 2  
; LENGTH: 1514  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-333-729A-2

Query Match 100.0%; Score 1100; DB 3; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCCCGGGCTCTGGGACCGCGCGGTACCTGCTCTACTGCCCCCTGCATNGGGCGC 60  
Db 1 ATGCCCGGGCTCTGGGACCGCGCGGTACCTGCTCTACTGCCCCCTGCATNGGGCGC 60  
QY 61 TTGGGAACACGAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGTAACCGT 120  
Db 61 TTGGGAACACGAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGTAACCGT 120  
QY 121 ACCTTGGCTGTCCTCTCTGGATTCAGTACAGCATCACAAAGCTCTCTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCTCTCTGGATTCAGTACAGCATCACAAAGCTCTCTTACCAACCTC 180  
QY 181 CATGTGCTTACCAAGACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTATC 240  
Db 181 CATGTGCTTACCAAGACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTATC 240  
QY 241 AGCTTGGAGGATTTCAATGGAGAGCTGGACCCACCCACCTGGCCCCCTGAGAAGCGGTG 300  
Db 241 AGCTTGGAGGATTTCAATGGAGAGCTGGACCCACCCACCTGGCCCCCTGAGAAGCGGTG 300  
QY 301 GCATACTGCTTGGAGTGGCAGCCAGCAAGCCAGATTAAGAGACGTGCCCCATGAAG 360  
Db 301 GCATACTGCTTGGAGTGGCAGCCAGCAAGCCAGATTAAGAGACGTGCCCCATGAAG 360  
QY 361 GAAGGAACCCCTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAAGTCGGAG 420  
Db 361 GAAGGAACCCCTTGGCCCATCTGGGATCAGTTTTCATGTGAGTTTCAAGTCGGAG 420  
QY 421 CTTTTTACAGGCAATTCCTTCACTGCTGCTTCTACAGAAACAATGGAGCCAGATTTTCT 480  
Db 421 CTTTTTACAGGCAATTCCTTCACTGCTGCTTCTACAGAAACAATGGAGCCAGATTTTCT 480  
QY 481 CCAAGGAACATCCGGTGTCTTGGCTTGGCCAGGAGCCCGCCAGTTCCTCCCTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTCTTGGCTTGGCCAGGAGCCCGCCAGTTCCTCCCTCTAGAA 540  
QY 541 GAACACAGGCCACTACAGAGTACATGTTGTTGTCAGAGAAATGTTGAAGACGGAGAG 600  
Db 541 GAACACAGGCCACTACAGAGTACATGTTGTTGTCAGAGAAATGTTGAAGACGGAGAG 600  
QY 601 GCCAGATTATGCCACCTTGTCCGGCCCTATGTGGGATTCATCTGGGATTTGGTCT 660  
Db 601 GCCAGATTATGCCACCTTGTCCGGCCCTATGTGGGATTCATCTGGGATTTGGTCT 660

QY 661 GACTGGAAGAACCCCTGTGCGCATGCTGAAGGACGGGACTGCGAGGCTCGCACTTATGSCC 720  
Db 661 GACTGGAAGAACCCCTGTGCGCATGCTGAAGGACGGGACTGCGAGGCTCGCACTTATGSCC 720  
QY 721 TCTCCGCAAGTGTGTGGGCTACAGCCGACAGCAAGCGGCCCCCTCACGATGATATGTGC 780  
Db 721 TCTCCGCAAGTGTGTGGGCTACAGCCGACAGCAAGCGGCCCCCTCACGATGATATGTGC 780  
QY 781 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTCTGGTGTGATGC 840  
Db 781 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTCTGGTGTGATGC 840  
QY 841 CAGTCGCTTACCTGCTACTGATTCGAGAGTTATGTCCTGAGCTCCAAACAGCTCTTC 900  
Db 841 CAGTCGCTTACCTGCTACTGATTCGAGAGTTATGTCCTGAGCTCCAAACAGCTCTTC 900  
QY 901 AAAGGAAAGTGAAGGTGGTGAAGCTTGAAGCTGAGGTGGCCCCAGGTGCACTGTACATC 960  
Db 901 AAAGGAAAGTGAAGGTGGTGAAGCTTGAAGCTGAGGTGGCCCCAGGTGCACTGTACATC 960  
QY 961 CTGCGCCAAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 961 CTGCGCCAAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
QY 1021 CGGAGAGCGGACCTCCAGGGGAGGCGCTCTTCTTTCGSCATGGACAGGCCCTTAAG 1080  
Db 1021 CGGAGAGCGGACCTCCAGGGGAGGCGCTCTTCTTTCGSCATGGACAGGCCCTTAAG 1080  
QY 1081 CTGCGGAGCAGGTTCTGATT 1100  
Db 1081 CTGCGGAGCAGGTTCTGATT 1100

## RESULT 5

US-08-978-741-4  
; Sequence 4, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1300 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear

US-08-978-741-4

Query Match 99.9%; Score 1098.4; DB 3; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTTACTGCTGATGGGGCGC 60  
DB 136 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTTACTGCTGATGGGGCGC 195

QY 61 TTGTGGGAACCGGCGGATCACTCTTGGGCTCTGGGATTTGGAAGCTGCTAAACCGT 120  
DB 196 TTGTGGGAACCGGCGGATCACTCTTGGGCTCTGGGATTTGGAAGCTGCTAAACCGT 255

QY 121 ACCTTGGGCTGCTCCCTCTTGGATTGAGTACAGCATCACAGCCCTCTTCAACCAACCTC 180  
DB 256 ACCTTGGGCTGCTCCCTCTTGGATTGAGTACAGCATCACAGCCCTCTTCAACCAACCTC 315

QY 181 CATGTGCTCTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTGATC 240  
DB 316 CATGTGCTCTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTGATC 375

QY 241 AGCTTGGAGGATTTTCATGAGNAGTGGCACCACCACTGCCCCCTGAGAGCGGGTG 300  
DB 376 AGCTTGGAGGATTTTCATGAGNAGTGGCACCACCACTGCCCCCTGAGAGCGGGTG 435

QY 301 GCATACTGCTTGGGTCGAGCGGACCCAGCGAGCCAGATTAAGAGAGCTGCCCATGAAG 360  
DB 436 GCATACTGCTTGGGTCGAGCGGACCCAGCGAGCCAGATTAAGAGAGCTGCCCATGAAG 495

QY 361 GAAGAAACCCCTTTGGGCCATTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
DB 496 GAAGAAACCCCTTTGGGCCATTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 555

QY 421 CTTTTTCAGGATTTCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 480  
DB 556 CTTTTTCAGGATTTCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 615

QY 481 CCAAGGAACATCCGCTGCTGCGGCTGCGAGGCGCCAGCCAGTTCCTCCGCTCTAGAA 540  
DB 616 CCAAGGAACATCCGCTGCTGCGGCTGCGAGGCGCCAGCCAGTTCCTCCGCTCTAGAA 675

QY 541 GAACACAGGCCACTACAGAAGTACATGTATGGTCAGACGAATGGTGAAGACGGGAG 600  
DB 676 GAACACAGGCCACTACAGAAGTACATGTATGGTCAGACGAATGGTGAAGACGGGAG 735

QY 601 GCCCAGATTCATGCCCATCTTGTGCGGCTTATGTGGGCAATTCATCTGCGCAATGGCTCT 660  
DB 736 GCCCAGATTCATGCCCATCTTGTGCGGCTTATGTGGGCAATTCATCTGCGCAATGGCTCT 795

QY 661 GACTGGAAGACGCTGTGCTGCTGAAGACGGGACTGCGAGGCTGCGACTTCATGCGC 720  
DB 796 GACTGGAAGACGCTGTGCTGCTGAAGACGGGACTGCGAGGCTGCGACTTCATGCGC 855

QY 721 TCTCCGCAAGTGTGGGTACAGCGCAGCAGCGGCGCCCTCACCAGTACATATGTC 780  
DB 856 TCTCCGCAAGTGTGGGTACAGCGCAGCAGCGGCGCCCTCACCAGTACATATGTC 915

QY 781 CTGCTTACCTGAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 840  
DB 916 CTGCTTACCTGAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 975

QY 841 CAGTCGGTCTAGTTGCTTACTGATTTCCAGAGATTATGTGCTGAGCTCCACAGCTCTTC 900  
DB 976 CAGTCGGTCTAGTTGCTTACTGATTTCCAGAGATTATGTGCTGAGCTCCACAGCTCTTC 1035

QY 901 AAAGGAAGGTGAAGGTGGTGAAGCTGAGGTGGCCAGGTCGACTCTGATCATC 960  
DB 1036 AAAGGAAGGTGAAGGTGGTGAAGCTGAGGTGGCCAGGTCGACTCTGATCATC 1095

QY 961 CTCGCCCAAGCGGACCACTTTATTGGCAATGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
DB 1096 CTCGCCCAAGCGGACCACTTTATTGGCAATGTGTCTCTCTTCACTGCTTTGTGAAG 1155

RESULT 6

US-09-333-729A-6  
; Sequence 6, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: F1041P1D1-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 6  
; LENGTH: 1300  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Plasmid insert.  
US-09-333-729A-6

Query Match 99.9%; Score 1098.4; DB 3; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTTACTGCTGATGGGGCGC 60  
DB 136 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTTACTGCTGATGGGGCGC 195

QY 61 TTGTGGGAACCGGCGGATCACTCTTGGGCTCTCTGGGATTTGGAAGCTGCTAAACCGT 120  
DB 196 TTGTGGGAACCGGCGGATCACTCTTGGGCTCTCTGGGATTTGGAAGCTGCTAAACCGT 255

QY 121 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCACAGCCCTCTTCAACCAACCTC 180  
DB 256 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCACAGCCCTCTTCAACCAACCTC 315

QY 181 CATGTGCTCTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTGATC 240  
DB 316 CATGTGCTCTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTGATC 375

QY 241 AGCTTGGAGGATTTTCATGGAAGCTGCGACCCACCACTGCGCCCTTGAGAGCGGGTG 300  
DB 376 AGCTTGGAGGATTTTCATGGAAGCTGCGACCCACCACTGCGCCCTTGAGAGCGGGTG 435

QY 301 GCATACTGCTTGGGTCGAGCGGACCCAGCGAGCCAGATTAAGAGAGCTGCCCATGAAG 360  
DB 436 GCATACTGCTTGGGTCGAGCGGACCCAGCGAGCCAGATTAAGAGAGCTGCCCATGAAG 495

QY 361 GAAGAAACCCCTTTGGGCCATTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
DB 496 GAAGAAACCCCTTTGGGCCATTTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 555

QY 421 CTTTTTCAGGATTTCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 480  
DB 556 CTTTTTCAGGATTTCTTTCAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 615

QY 481 CCAAGGAACATCCGCTGCTGCGGCTGCGAGGCGCCAGCCAGTTCCTCCGCTCTAGAA 540  
DB 616 CCAAGGAACATCCGCTGCTGCGGCTGCGAGGCGCCAGCCAGTTCCTCCGCTCTAGAA 675

QY 541 GAACACAGGCCACTACAGAAGTACATGTATGGTCAGACGAATGGTGAAGACGGGAG 600







Db 607 TTCACTAGAGAACTCAAGCTGACGCGCTACTTTCAGTGGAGTCAACGGTATAGGGAAG 666  
Qy 593 CGGAGAGGCCAGATTATGCGCCACCTTGTCCGCGC---CTATGTGGCATTATCTGCG 649  
Db 667 CATCTAAGGATTTATCCGAGAGCAGTGTGCTCGGGGTGCCTTTTGGGCAATTCATCTGC 726  
Qy 650 GCAATGGCTCTGACTGGAAGAACGCTGTGCCATCTGCAAGGACGG-----GACTG 700  
Db 727 GCAACGGTATCGATTGGGTGAGAGCTGTGAGCAGTCAAGGATAGCCAGCATCTGTGTTG 786  
Qy 701 CAGGCTCGCACTTATGSCCTCTCCGCAAGTGTGCGGTACAGCCGACAGCAGCGGCC 760  
Db 787 CCTCGCCGAGTGTCTGGGCTATAAAATGAACGTGGTGCACTCTACCGGAGCTCTGCA 846  
Qy 761 CCCTCAGCATGACTATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 820  
Db 847 TGCCCTCCAAGGAGCGCATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTCGCCCAA 906  
Qy 821 GGGTGGGTGCTGGATGCCAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 880  
Db 907 CTCAGCCGACACCAAGATCAATCAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966  
Qy 881 CTGAGCTCCACAGCTCTTCAAGGGAAG---TGAAGGTGGTGGAGCCTGAAGCCTGAGG 937  
Db 967 GTGAACCTAAACACGCGCTTGTAGTGGCATGGGCTGCTGCTGCTGCTGCTGCTGCTG 1026  
Qy 938 TGGCCAGGTGCACTGTATCATCTCTCGGCAAGCCGACCTTTATTTGGCACTGTGCT 997  
Db 1027 ATCTTACTGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1086  
Qy 998 CTTCTTCACTGCTCTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGCTCTCTTCT 1057  
Db 1087 CTTCTTACTGCGCATCTGTAAAGGGAACGAGATGTGACGGTTTTCATGCTACTTCT 1146  
Qy 1058 TGGCATGCGCAGG 1071  
Db 1147 GGGGATTCGCCAAG 1160

RESULT 11  
US-09-270-767-30507  
; Sequence 30507, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 30507  
; LENGTH: 722  
; TYPE: DNA  
; ORGANISM: *Drosophila melanogaster*  
US-09-270-767-30507

Query Match 5.7%; Score 62.2; DB 4; Length 722;  
Best Local Similarity 49.5%; Pred. No. 2.2e-08;  
Matches 225; Conservative 0; Mismatches 218; Indels 12; Gaps 2;  
Qy 629 CCTATGTGGCATTTATCTGGCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 688  
Db 108 CTTTGTGGCATTTATCTGGCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 167  
Qy 689 AGGACGG-----GACTGAGGCTCGCACTTCACTGCTCTCCGAGTGTGGGCT 739  
Db 168 AGGATAGCAGCATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 227  
Qy 740 ACAGCCGAGCAGCAGCGGCCCTTCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 799  
Db 228 CACTCTACCGGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 287

Qy 800 TCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGATGCCAGTCCGCTTACGTTGCTA 859  
Db 288 CCATTAGAAGCTGCGCCAAACTCAGCCGAGACAGAAATCAATCAGTTTCTGTTGGGT 347  
Qy 860 CTGATTCGAGAGTATGTGCTGAGCTTCCCAACAGCTCTTCAAAGGGAAG---TGAAGG 916  
Db 348 CAGACTCCAATCACATGATTGGTGAACCTAAACAGGCTTGTAGTCGATGGSCATCAGTG 407  
Qy 917 TGGTGAAGCTGAAGCCTGAGGTGGCCAGGTGCTGACCTGTATCTCTCGGCCAAGCGGACC 976  
Db 408 TGCAACAGCTCAGCGAGGATGATCCTTACCTGGACTTGGCCATTTCTCGACAGTCCGAAC 467  
Qy 977 ACTTTATTTGGCACTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1036  
Db 468 ACTTTATCGCACTGTATATCTCTTACTCGCATTCGTAAGGGAACAGATGTGC 527  
Qy 1037 AGGGAGGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1071  
Db 528 ACGGTTTTCCATCGTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 562

RESULT 12  
US-09-270-767-28905/c  
; Sequence 28905, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 28905  
; LENGTH: 998  
; TYPE: DNA  
; ORGANISM: *Drosophila melanogaster*  
US-09-270-767-28905

Query Match 5.3%; Score 58.6; DB 4; Length 998;  
Best Local Similarity 60.2%; Pred. No. 3.2e-07;  
Matches 97; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
Qy 911 TGAAGTGGTGGAGCTGAGCCTGAGCTGAGGTGGCCAGGTGCTGATCTCTCGGCCAAG 970  
Db 896 TCAGTGTGCAACAGCTGACGAGGATGATCTTACCTGGACTTGGCCATTTCTCGACAGT 837  
Qy 971 CCGACCACTTTATTTGGCACTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1030  
Db 836 CGAACCACTTTATCGGCACTGTATATCTTACTCGGCATTCGTAAGGGAACGAG 777  
Qy 1031 ACCTCCAGGGAGGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1071  
Db 776 ATGTGACGCTTTTCCATCGTACTTCTTGGGATTCGCCAAG 736

RESULT 13  
US-09-270-767-13022/c  
; Sequence 13022, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 13022  
; LENGTH: 2034  
; TYPE: DNA  
; ORGANISM: *Drosophila melanogaster*

US-09-270-767-13022

```
Query Match          5.3%; Score 58.6; DB 4; Length 2034;
Best Local Similarity 60.2%; Pred. No. 4.8e-07;
Matches 97; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 911 TGAAGGTGCTGAGCTGCAAGCTGAGGTGGCCAGGTGACATCTTCGGCCCAAG 970
Db 1932 TCAGTGTGCAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATCTCGGACAGT 1873

Qy 971 CGGACCACTTATGCGAACTGTCTCTCTCTTCACTGCTTGTGAAGCGGAGCGGG 1030
Db 1872 CGAACCACTTATGCGCACTGTATCTCTTACTCGGCATTCGTAAGGGAACGAG 1813

Qy 1031 ACCTCCAGGGAGGCGCTCTTCTTCTTCGGCATGGAACAG 1071
Db 1812 ATGTGCAGGTTTTCCATGCTGATCTTGGGGATTCCCCCAAG 1772
```

RESULT 14

```
US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12127
; LENGTH: 19227
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12127
```

```
Query Match          5.1%; Score 56; DB 4; Length 19227;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCCGGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 56
Db 1718 ATGCCCCGGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 1663
```

RESULT 15

```
US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16285
```

```
; LENGTH: 19228
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-16285

Query Match          5.1%; Score 56; DB 4; Length 19228;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCCGGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 56
Db 1718 ATGCCCCGGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 1663

Search completed: November 9, 2005, 10:46:08
Job time : 202.044 secs
```